FOOD AND DRUG ADMINISTRATION

PUBLIC WORKSHOP

NEXT-GENERATION SEQUENCING-BASED ONCOLOGY PANELS

Thursday, February 25, 2016

8:31 a.m. to 4:24 p.m.

FDA White Oak Campus

10903 New Hampshire Avenue

Building 31 Conference Center

Silver Spring, Maryland

Meeting Roster

Dara Aisner

Eliezer Van Allen

Michael Berger

Gideon Blumenthal

Joshua Deignan

Jennifer Dickey

Dane Dickson

David Eberhard

Soma Ghosh

Madhuri Hegde

Yun-Fu Hu

Robert Klees

Greta Kreuz

Shashi Kulkarni

Eunice Lee

You Li

Sharon Liang

Rajyalakshmi Luthra

Elizabeth Mansfield

Anand Pathak

John Pfeifer

Reena Philip
Donna Roscoe
Michael Rossi
Aaron Schetter
Jeffrey Sklar
Apostolia-Maria Tsimberidou
Abraham Tzou

C O N T E N T S

C O N T E N T S (continued)

P R O C E E D I N G S

(9:03 a.m.)

DR. DICKEY: Well, hello, and welcome to the FDA public workshop on Next Generation Sequencing Oncology Panels. We are very excited to have you here for what we hope will be a lively and informative discussion.

Just a few housekeeping details. This public workshop is being webcast. The address is here on the slide. Within the next week, after the presentation, the archive of the webcast will be available on the website. Please set your phones, computers, and Blackberrys to silent mode. You can access WiFi in the guest room using the code publicaccess, I believe, though there has been some trouble with that.

Food and beverages are available for purchase at the kiosk in the registration lobby during breaks and lunch. I highly recommend pre-purchasing box lunch during break 1 to speed the line at lunch time. Links to the meeting transcripts will be posted six to eight weeks after the meeting.

So with that said, we can get started with

our introduction and welcome. I'd like to introduce Elizabeth Mansfield. She's the deputy office director for personalized medicine here at FDA.

Welcome and Introduction - Elizabeth Mansfield

DR. MANSFIELD: Thanks, Jennifer, and thank you all for coming today. I'm really glad that we're able to get extra room. I understand that there are still a lot of people trying to get through security, but we need to move along with our meeting, so we'll get started now.

We're really happy that you're here. We're really interested in hearing what you have to say to us about the topics that we've laid out. I can assure you we've done a lot of thinking on it ourselves, but we're anxious for external input.

As you all know, oncology panels that use NGS are already widely used clinically and are and will be an important tool in precision medicine as we move forward. However, this meeting is not about the Precision Medicine Initiative. This is about addressing NGS oncology panels under the regulatory paradigm that we have in place today. The Precision

Medicine Initiative is moving along, but we're not quite this advanced yet; so just to dispel any confusion about what we're talking about.

These panels, I think as you all know, are effectively multiplex assays that can interrogate a huge number of analytes, some of which may have companion diagnostic status and some of which may not yet. We expect them to be used, and they probably are being used, for treatment selection, especially in oncology. We know that they are important in this area for cancer patients, especially in the areas of lung cancer right now, where a single patient may need to have multiple different tests, and there's simply not enough tissue available to do individual tests, and probably not enough time to do all the tests separately.

So we think panels are a really good way to move forward in oncology where you can get all the information you need out of one test. So I just want to start off saying FDA actually supports this idea.

Another caveat here, the focus of this meeting is how we will address the regulatory aspects

of NGS oncology panels, some of which we are aware may be laboratory developed tests. This meeting is not about whether FDA should regulate laboratory developed tests. I realize that people may have strong opinions one way or the other. That's not a topic of discussion today. So I would ask people to try not to distract from the meeting with that particular subject. We're more looking at scientific aspects for validation and so on.

At this meeting, we're going to introduce some concepts and questions around the intended use of an NGS oncology panel; how should we deal with reporting of variants that don't have specific clinical validity yet where we think that they may be important, in particular for therapeutic uses; how do you actually address what turns into a follow-on companion diagnostic or a companion diagnostic that has the same intended use as a test that we've already approved; is there a way that we can allow rapid modification of these tests because we realize that the science moves fast, clinical trials move pretty fast, and we want the test to be able to keep up with that, but we need to

figure out exactly how to address that. So we'll be discussing that.

We would like to discuss whether and how a representative subset of analytical variants for validation would be sufficient to analytically validate a larger panel. We realize, in many cases, people are looking at whole genes. They're sequencing whole genes. They're not just targeting specific mutations, and how do we actually address the analytical validation for things that you weren't necessarily looking for in the first place.

The meeting format, we're going to have three moderated panel discussions with some excellent panelists. Then that's going to be followed up by a comment period. The first panel will be the pre-analytical challenges towards these tests, which are very interesting in terms of how you're getting the tissue, what's the nature of the tissue that you're getting and so on.

The second panel will be around analytical challenges or how you establish your analytical performance, especially with analytes that have

different meaning when they're offered within the same tests, and what are the clinical challenges, how do you establish clinical validity for companion diagnostics, for follow-on companion diagnostics, and for other variants that may be on the panel.

We did post a discussion paper, which I hope that you all were able to access and read prior to the meeting. If you haven't, there's still a chance if you have WiFi access. We also have a docket open that you can make comments to. And we greatly appreciate comments, especially comments that specifically address the discussion that we're having today. We will have that open until March 28th, so you have about a month to get your comments, and we, again, appreciate all types of feedback.

So I leave you in the very capable hands of the team that's been working on this for a couple of years now and hope that you all have a productive and meaningful day, and that you come away from this meeting like we hope, with more information in your pocket. So again, thank you all for coming, and enjoy the meeting.

Now, I guess I turn it over to someone, but I don't know who. Reena. Dr. Reena Philip is the division director for the Division of Molecular Genetics and Pathology, and that is the group that will be handling the regulations of these types of tests.

Overview of Goals and Background - Reena Philip

DR. PHILIP: Thank you, Liz, for the great introduction.

Thank you all for coming to this exciting workshop, and I want to especially thank the panelists for accepting our invitation. My task is to go over the overview of the meeting goals.

The outline of my talks is as follows. I'll provide a brief background about companion diagnostics and the follow-on companion diagnostics because those two concepts will be discussed throughout this workshop. I'll go over the scope of the workshop. We'll present a hypothetical case, and we'll briefly discuss the workshop discussion topics.

As Liz mentioned, the first panel will be pre-analytical and quality metric approaches. The second is about the analytical validation, and the

third one is about the clinical. But I will go over the potential general intended use, which is actually put out in the discussion paper, and you probably already took a look at it. Please keep that in mind when we are discussing the three sessions.

I may be preaching to the choir about the importance of companion diagnostics in personalized medicine, but having accurate reproducible and clinically useful tests are so important in personalized medicine. Companion diagnostics are the tests that provide information that is essential for the safe and effective use of the corresponding drug or a biological product.

We had finalized the guidance on companion diagnostics in 2014 that defines the companion diagnostic device and various scenarios for use. It also describes our policies for approval and labeling, and it recommends the contemporaneous approval of both device and the drug.

For every premarket review, the two critical components of the review are the analytical validation and the clinical validation. The analytical validation

for companion diagnostic devices is conducted with clinical specimens from the intended-use population. Of course, there are exceptions for rare mutations.

This analytical validation – includes for example, accuracy, reproducibility, and the other studies we require; they are usually obtained with attention to the clinical decision point. Also, we make sure the studies are aligned with the technology that we review. If it's IHC, we ask for specific validation studies. If it's a molecular assay, there are some studies that we require but maybe not for IHC.

The clinical validation is supported by the results from the drug trial. The companion diagnostic or CTA, clinical trial assay, may be used to test specimens and identify patients. And if a CTA is used, we ask for the bridging studies that demonstrates the drug efficacy using that particular companion diagnostic.

There may be some subsequent devices that may be coming after one particular companion diagnostic is approved for a particular intended use with that therapeutic indication; we are calling these companion

diagnostics “follow-on companion diagnostics.” They have the same intended use and therapeutic indication as the originally approved one. The first one is probably called the original, and the follow-ons are the ones which come right after that.

Since the follow-on has the same intended use and therapeutic indication, it should consistently and accurately select the same intended-use population as the original one, and it should demonstrate the same or comparable level of the analytical and clinical performance for the specific intended use which they claim.

So far, all the companion diagnostics that we have approved, assess a single analyte or prespecified mutations which is associated with that therapeutic indication. You all probably know the trend these days. NGS tumor panels are increasingly used, and that's because they can interrogate a patient's tumor specimen for numerous biomarkers. That introduces challenge to the current companion diagnostics regulatory paradigm, and this workshop has actually been called to discuss those regulatory paradigms, the

complexities, and you will understand why we face difficulties with this paradigm in my subsequent slides.

I'll go over the scope of this workshop. It's to get input from external stakeholders on the analytical performance of all these panels, because these panels will include variants that are intended to be used as companion diagnostics and also includes the other variant that may be used for alternative therapeutic management of patients who have already been considered for all appropriate therapies. The second is also to get information on the clinical data that's needed to support the follow-on companion diagnostic devices.

We're also requesting input on strategies for establishing performance characteristics for the rare variants. Again, we are seeking input on the claims for follow-ons and also the post-approval assay modifications. That's something actually we're really interested to hear because we're thinking of some flexible regulatory paradigm for post-approval assay modifications. Your input on a post-approval assay

modification paradigm is something we are interested to hear.

I just want to reiterate what Liz said earlier. This is for the targeted NGS-based oncology panels that are actively marketed by the manufacturers, not focusing on LDTs today. It's just to get your input on what should be in the labeling so there is truth in labeling about what is being marketed and what the limitations of the assay are, to make sure there is adequate representation of panel performance so when a user, when a lab gets that test, they can decide on how and when to use that panel.

Again, the scope of this workshop is limited to targeted NGS-based oncology panels for human genomic DNA/RNA that are intended to be used as companion diagnostic devices for the clinical management of previously diagnosed oncology patients. It's also for the panels that can be used for alternative therapeutic management of patients who have already been considered for all appropriate therapies.

Today's workshop does not apply to subjects that have not been diagnosed with cancer. We are not

going to talk about circulating tumor DNA testing today. I just want to say there is a workshop on liquid biopsy July 19th. I hope the date is set. There could be some changes in the date, but there is a workshop planned for the liquid biopsy. Today's workshop will not address IVDMIAs using NGS, or genome sequencing, or exome sequencing, or for carrier screening. And we're not going to talk about the quality of the database because I know some people are interested in that.

I just want to go over the NGS-based oncology panel workflow when we look at an NGS-based oncology panel. It starts with specimens, then there is nucleic acid preparation, library preparation, sequencing, base calling, alignment, mapping, variant calling, annotation, interpretation, and reporting. So that's the NGS-based oncology panel workflow, which we look into when we're looking at a device.

Here's an example. This could be a 10-gene NGS-based targeted NGS-based oncology panel. Specimen source could be FFPE or fresh frozen for a solid tumor. It could be whole blood if it's a hematological tumor.

Analyte type could be DNA or could be RNA. There may be five genes that have the companion diagnostics claim; five maybe non-companion diagnostics claim. The alterations could then be all different categories: SNVs, insertions, fusions, translocations, gene amplifications; and then, of course, there is the genomic context, which could be simple or complex.

Although it may be pan-cancer NGS-based oncology panel, the companion diagnostics may have a specific indication, maybe only in colon cancer or in lung cancer. This could be an example of a submission or device that has a 10-gene, NGS-based oncology panel.

When you're looking at a device, you're considering the entire test system validation, from specimen collection, sample preparation, down to all the steps I mentioned earlier, up to the generation of the result report. The validation studies should be designed to demonstrate the performance characteristics of the device within the context of the intended-use population.

Of course, there are a lot of challenges, which we have already encountered. Some of them are

listed here but doesn't include everything. First of all, what genes and associated variants should be included in the panel, like if it's a pan-cancer claim, how can you qualify that particular gene or variant be included?

What are the limitations in reporting? I think Liz already mentioned about that. What are the units of validation? What's the most difficult unit we should be validating? Is it somatic or germline? Should we look at germline when looking at the somatic? Do we have to look into the matched blood if we have to actually compare it to germline? These are just a few challenges I listed, and I'm sure this workshop will talk about the different challenges we have encountered during the pre-submissions we have had with different sponsors.

I just want to briefly go over the workshop discussion topics. It's already in the discussion paper. You probably have read it, but I will review it for those who haven't read the discussion paper. Before I go over the topics that are going to be the panel, I want to talk about the potential general

intended use. Of course, there may be tweaks to it, but this is just our first thoughts on what could be a potential intended use for a targeted NGS-based oncology panel.

The device is a qualitative, in vitro diagnostic test that uses high throughput, parallel sequencing technology intended to detect sequence variations using whatever the particular instrument name. The device is indicated as an aid in characterizing sequence variations in X number of genes, maybe 10 genes, on DNA or RNA, and isolated from the particular specimen type, which may be FFPE, it may be blood, or whatever the specimen type is.

There will be a specific diagnostics claim versus a non-companion diagnostics claim. The companion diagnostics claim will be, as it is indicated, a companion diagnostic to aid in selecting oncology patients for treatment with the targeted therapies listed in this table 1, and that's in accordance with the approved therapeutic product labeling. You can see that table 1 has gene, variants, tissue types, targeted therapies.

There is a table 2. Results other than those listed in table 1 are only intended for patients who have already been considered for all appropriate therapies. Safe and effective use has not been established for selecting therapy using this device for the variants and the associated tissue types not listed in table 1.

Table 1 is only for companion diagnostics device, and table 2 will have the variants that demonstrated analytical performance characteristics and there is a disclaimer that it's not intended for standalone diagnostic purposes, screening, monitoring, risk assessment, or prognosis.

There are a lot of questions we have thought about regarding the intended use statement, and those questions won't be discussed today, but you can actually provide your input to the docket and provide answers to these questions. In general, that's the intended-use, to capture the necessary elements to be able to use and interpret the targeted NGS-based oncology panel.

That's the main question, and then some other

questions like should the tissue types be included in table 1 because table 2 had the analytical performance characteristics of the ones which didn't have the companion diagnostics claim, but they probably provided samples from particular tissues. So should that be reported in table 2?

What level of analytical validity should be established for variants reported by the assay not included in table 1 or table 2? There will be other variants that will be reported by the assay, but that's not in table 1 or table 2. So what kind of analytical validity should be established for those?

For clinical validity, what clinical validity should be established for any genes reported by the assay? I think I already mentioned about this earlier. How do you make sure you get into that NGS panel, and what level of validity should be established for getting into it? Would evidence of a clinical trial be sufficient?

What types of warnings or disclaimers should be included for variants reported by the assay but not included in table 1 or table 2? There may be something

which should not get reported, so what type of warnings should be included for those? What warnings should be included for de novo variant reporting as opposed to predefined variant reporting?

The first panel will go over the pre-analytical and quality metric approaches. As I mentioned earlier, traditionally, we have asked for clinical specimens from all specimen types that are specified in the intended-use statement. But it's not clear whether information about each processing parameter across each tissue type is needed for this. These are the challenges and the complexities and the regulatory paradigm for NGS, number one. Do we need information about each tissue type that is actually listed in the NGS oncology panel?

So we are seeking input on whether there are suitable pre-analytical tests or with a representative set of sample types or QC metrics that may be used instead of requiring all sample types and processing parameters to demonstrate robustness for that particular targeted NGS-based oncology panel. There are some questions that, Aaron will go over in his

session but just a few I have listed here.

Are there pre-analytical steps that are most critical? Are there tumor types that are more challenging and in what processing context? What could be the appropriate level of validation needed to support the different claims? The second session will go over the analytical validation bioinformatics and the potential for a flexible regulatory paradigm for post-approval assay modifications.

As we saw earlier, this NGS-based oncology panels will report variants over a spectrum of ranges from companion diagnostic indications to variants of uncertain significance. We are seeking input on the appropriate level of analytical validity that should be established and demonstrated for those variants.

Donna will go over the questions during the analytical session. Some of the questions I've listed here. Should the number of variants reported by the panel determine whether a representative variant approach is acceptable? Are there parameters that are most important to capture in a representative variant set? Are there differences in the sequencing platform

that would impact selection of a representative variant set? Once analytical validity has been satisfactorily established for a set of variants, are there requirements or controls that can be in place to add, subtract, or substitute variants as post-approval assay modifications?

The last panel before the public comments is on the clinical validation, and Dr. Abe Tzou will go over the questions for those. But it's mainly -- as you saw earlier, the companion diagnostics indication is in the intended use, but there's also non-companion diagnostic indication.

The companion diagnostic, it's easy. If it's intended to guide therapy and you have a clinical validation, that will go on table 1. The non-companion diagnostic that actually demonstrated the analytical performance, will go in table 2, but there will be others reported that's not in table 1 or table 2. Some of the questions will be what are the key considerations for evidence that would not be sufficient for follow-on companion diagnostic claim? What are the appropriate expectations for routine

reporting of genes or variants without established CoDx claims? What are the warnings or disclaimers that should be considered around issues of panel comprehensiveness? And what level of validation should be needed to move a variant from table 2 to table 1 when new targeted therapies are approved?

You may say I'm already in table 2. I don't need to actually do anything to get in table 1, but what kind of validation should be needed. And what are the warnings or disclaimers that should be considered around de novo reporting?

The video archive of this workshop will be posted next week, and discussion materials have already been posted on the website. You've probably already read it; you know everything about it. Please comment on the discussion materials prior to March 28th. We are really looking for input, and comments can be made to the docket or you can send via email. And I just want to reference the FDA website and companion diagnostics, and also the guidance I mentioned, the FDA companion diagnostic guidance.

With that, I'll turn it over to Dr. Aaron

Schetter for the first panel.

(Applause.)

Panel 1 - Aaron Schetter

DR. SCHETTER: I would welcome the Panel 1 members to come up to sit. Again, thanks, everyone, for coming today, taking time out of your schedules to come. We really value the expertise that you can help provide and guide us in making the decisions we have to make about the NGS oncopanels.

I have to say that we have a great panel here today. The panelists are all experts in molecular pathology and have extensive expertise in the clinical use of NGS oncopanels. FDA is seeking input from this panel on the pre-analytical quality metrics approaches for NGS oncopanels.

The format is going to be we'll start off with I'll give a brief introduction. We'll have two 10-minute talks from Dr. John Pfeifer and Dr. Dara Aisner, at which point then we'll start going through and open it up for the questions, which we'll then have panel discussions. And we hope to have 10 minutes or so at the end to open it up for public questions.

Pre-analytical variability in tissue handling and processing is expected in laboratory use, and manufacturers are going to have very little control over the tissue-handling steps. Manufacturers of the NGS oncopanels will be expected to provide data that indicate how their assay performs under a reasonable number of these pre-analytical conditions that they're expected to see as part of routine laboratory use.

To frame our discussion, currently labs that employ NGS-based oncopanels use research-use only reagents, and therefore, the laboratories have to perform full validation of all of these assays in their labs in order to show it can be used for the intended use.

FDA seeks input from the laboratory perspective about the types of data that should be provided by a manufacturer to the FDA and an IVD labeled assay to demonstrate a robust performance of the assay across the variety of conditions in a manner that would alleviate the need for a lab to perform full validation of the assay once the assay is approved. Rather, a lab could take the approved assay and perform

verification instead of full validation, which could lead to less burden on the laboratories.

With that, I'll turn it over to Dr. John Pfeifer, who's the director of molecular pathology at the laboratory at the Washington University School of Medicine.

Presentation - John Pfeifer

DR. PFEIFER: First of all, thank you very much for inviting me to speak here this morning. Here's the workshop and Panel 1, Pre-analytical and Quality Metric Approaches, and the required disclosures slide. As I always say in a meeting like this, the most important part of that slide is what's on the right.

I'm a faculty member at Washington University School of Medicine in the Department of Pathology, and we have a next-generation sequencing lab, which we call Genomics and Pathology Services. What's important about all of that is if living and working in an academic tertiary care urban medical center doesn't bias your view of the world, nothing will.

So this is really -- when you come to a

meeting like this and people start talking about next-generation sequencing, this is the slide that everybody puts up to talk about the workflow. And what's interesting about this first panel, and I think most of what we're talking about today, is step number 1. And we're not even to step number 1. The purpose of what we're talking about is how do you even get to step number 1. And that's a key feature because most, or many, variables that have a huge impact on the results that you're going to get actually occur before people start talking about the workflow for next-generation sequencing.

So what I'm going to do today is I'm going to talk about, in general terms, what those pre-analytic steps are to sort of tee up the conversation for Panel 1. Dara in her talk is going to go through a number of these in more detail, and we'll certainly talk about them more in the panel. But I want to just make sure everybody's on the same page and set the landscape for what these key parts are, and in general terms mention where the uncertainty can arise, where the quality issues can arise, and try and tie that to what we know

based on evidence-based medicine, what we know from the literature, so that our conversations are grounded on what we know to be facts rather than things that we just worry about at night when we're falling asleep.

So in general -- and we can break these down differently. But for the purposes of my talk, we're going to talk about four parts of that pre-analytic before you even get to step 1. The first thing is the gross processing of the tissue. You get a different result depending on where it is that you sample the tumor. We need to be aware of that.

Second of all, tissue processing. People worry a lot about formalin fixation, but, news flash, there are similar sorts of chemical changes that occur even in fresh tissue that we need to be aware of and how big are those.

The third thing is this whole concept of histopathologic review in tumor enrichment. Then the last point is this business about DNA extraction. People worry about quality of DNA, and people worry about amount of DNA. But there's little thing called library complexity, which can have a huge impact on

this, which is often ignored.

Tumor sample. This is the first part of that quality metric, and what we know now -- and this goes back to some papers that are probably five years old now, and one of them is an early one by Vogelstein's group -- is that there is a lot of heterogeneity within a tumor. It's due to clonal evolution. And we need to be aware of the significance of this heterogeneity when we start talking about these quality metrics.

What we know -- and this is from a study that looked at pancreatic cancer -- is depending on where you sample in a primary tumor, you will find or not find important biologic mutations in drivers. Now, the importance of that is two labs can look at the same tissue and differ in terms of what mutations they find in important driver genes not because of differences in the quality of their assay; simply because they're looking at different parts in a tumor.

So we need to be aware in all of this that the same tumor, that the resection of a tumor -- and the same thing is true, incidentally, for metastases. The metastases don't all originate at the same point in

a tumor's evolution. There at that bottom panel, they drop off at different points. If you're going to sample the primary tumor, you will get different results depending on where you sample the tumor. And if you go for metastatic disease, depending on which metastasis you sample or at which tissue site, you will also get a different answer.

So as we talk about assay validation, as we talk about the way we set as panels, we need to understand that out of the gate, that if we sample the same tumor in different places even within the same lab, you're going to get a different answer. So generally, this is an important feature because, as I keep saying, these are not trivial passenger mutations. These are often in the oncogenic drivers.

Now, having said all of that, we then come to the point where we talk about fixation. And there's a lot of concern about formalin fixation. What I want to basically do is just sort of demystify this a little bit so that we can start having a conversation about what formalin fixation does and what we need to be concerned about that's based on the data rather than a

lot of impressions about what's going on.

Formalin fixation is a chemical process. The next slide, I'm going to show you what it does. But here's a very important piece of information that we need to know. We did a study where we compared 17 fresh-frozen and lung adenocarcinomas. So we took some stuff from the tissue bank, and we took the corresponding wet tissue.

We ran our panel on it. We sequenced to a very high depth of coverage, and we found that while the distribution was slightly different, the mean of the fragment size was not different. Important. And the most important part is the panel at the bottom. There is more variability in your assay in terms of coverage, based on GC content between different places in individual genes, than there is between formalin fixation or fresh tissue. And that's a very important point to recognize.

At the bottom, there can be two or three or four or five-fold differences in your coverage, in the coverage of the assay you're performing, just due to GC content than there are due to the fact that the tissue

may or may not have been formalin fixed. So we have to keep that sort of thing in perspective.

More important, we need to make sure that we recognize what formalin fixation is actually doing. Formalin fixation actually causes a number of different changes. It leads to deamination, oxidation, cyclic-based derivative formation, and these methylene crosslinks.

Now, there's this view that what formalin fixation does is cause DNA degradation. I guess it does that. But primarily what formalin fixation does is it forms these crosslinks between the bases in DNA and RNA pyrimidines, purines, purines-purines, actually between nucleic acids and proteins. It's amine groups that actually -- it's the site that these crosslinks are formed.

Oftentimes what's viewed as DNA degradation is actually crosslinks that prevent the enyzmes that are involved in polymerizing DNA in the amplification steps of library preparation, so the DNA may be intact, but we just can't amplify it. So at the end of the day, it looks like it's degraded. Now, that's an

important point to recognize because some of the differences between formalin fixed tissue and fresh tissue aren't due to degradation, it's just we can't actually get into the DNA.

On the left-hand point, there's a very interesting comment. What we did in this study is we sequenced a very high depth of coverage. We looked at the errors due to formalin fixed, due to deamination, or oxidation, by base type and by local sequence context. And what we found is it's true that there are differences between formalin fixation, especially at the CpG dinucleotides that are sites of methylation. That's an important finding.

But the interesting comment there is how they rate. The rate is more common than in fresh tissue. It's not that formalin fixation introduces or causes changes that are unknown or unseen in fresh tissue, it's just that they occurred at a different rate. Many times, the rate isn't even two-fold difference if that. Sometimes the rate may be four-fold difference, but we're not really admitting the reality of all of this if we think that formalin fixation is causing changes

that aren't already there in fresh tissue.

At the end of the day, what we find in all of this is that there is a high degree of concordance between formalin fixed and frozen tissue. Only in about 1 in 10,000 bases is there a difference. But more importantly, there's a difference there that's probably a hundred-fold higher in difference between where you sample the tumor, the intratumoral heterogeneity.

So while there is a 100 percent concordance between orthogonal validation mutations, you find this intrinsic difference between fresh and formalin fixation is not due to the assay is formalin fixation, but it's primarily due to intratumoral heterogeneity, so we need to make sure we're paying attention.

Now, we also did one more experiment about this, which for the purposes of this we'll call the garbage DNA experiment, where we fix DNA for too long in formalin, for over a week. And what we found is there was no major difference in the percentage of mapped reads or on-target reads. Again, with increasing crosslinking, all we did was prevent

ourselves from polymerizing a subset of the DNA.

What was important, though, is we found increased coverage variability and a low number of unique reads because essentially what we've done is we've limited the population of molecules that were un-crosslinked that we could draw from. And that's a very important point because what it says is, is that the DNA that you do get from formalin fixed tissue doesn't have necessarily a high rate of mutations, it's just that it's hard; it's in lower quantity.

Now, just a few more slides to make a couple more points. I talked a lot about what we worry about on formalin fixation. Remember, there's a whole group of specimens out there that we don't even consider, we worry about. And these are not biopsy specimens or excision specimens, but cytology specimens, fine-needle aspiration specimens. And those are primarily fixed in ethanol and methanol, so we need to also worry about what chemical changes might be produced by different fixatives.

Then I want to briefly touch on this topic that Dara's going to mention a little more, and this

has to do with tumor sample variability enrichment.

Here's a low-power review of a cellular tumor. This probably has way over 50 percent cellularity. Many labs would simply look at this, make a number of sections, do no microdissection and put it into their assay. In our laboratory, we go in there. These are essentially 1 millimeter core biopsy, and each one of those biopsies has probably way over 95 percent cellularity.

Now, think about what this means if a laboratory has a minimum cutoff of tumor cellularity of 20 percent tumor cellularity to take a scroll and another laboratory says that's good, it's 20 percent cellularity, but we're actually going to enrich. What that means is suppose that it's a 20 percent cutoff and mutation is present in 20 percent of the tumor cells, actually 40 percent of the tumor cells, but present in a heterozygous state?

So the first lab, if they have a cutoff of 5 percent, are going to say, whoa, it's a 4 percent, we're not going to report it, where another lab like ours, it would be up around 40 or 50 percent. So two

labs with the same specimen with different workflows would come up with the different answer not because they're bioinformatics wasn't right, not because their assay wasn't validated, but simply because they were enriching or not enriching based on the tumor specimen that they were getting.

This actually drives home this point that Dara is going to present some details about, this histopathologic review. This is not some arcane point. It actually can impact assay sensitivity by four- or five-fold. And then the little secret in all of this that people don't pay attention to -- and I know I'm over, but just give me one or two more minutes -- is this idea about the DNA quality and sample requirement.

Yes, it's important to have a minimum amount of DNA or RNA for your assay, and laboratories will validate that, whether it's 5 nanograms or 10 nanograms. And yes, laboratories will look at the quality of their DNA. But as the graph on the left-hand side shows, you can get a certain amount of DNA from any amount of DNA input based on how many cycles of PCR you're willing to do. And what that

graph there on the upper left-hand panel shows is, even with several log-fold differences in the amount of input DNA, you can produce essentially the same amount of DNA for your assay.

So we're in a situation where one laboratory could start with a low tumor cellularity, low viability sample and produce enough nucleic acid for their assay, but essentially what they've produced is a very low complexity library. Library complexity is very easy to measure with certain hybrid capture-based assays, but for amplification based assays, it's harder to measure. And the point here is now we're no longer talking about differences, the four- or five-fold differences in sensitivity based on histopathologic review. We're talking on log orders of differences based on library complexity.

The problem here, as labs will say, we have a 5 or 10 nanogram input requirement, but they may be starting -- depending on how they do that library preparation, they may have log order differences in complexity. There's a much different information contact between 10 cells, 10 genome equivalents, than

there is from 1,000 cells and 1,000 genome equivalents.

Just to comment about when you start doing all this amplification, you increase the risk of contamination, which raises provenance issues. And then you get to the point typically that most people -- that's point number 1 on your diagram. So we've talked about all this stuff, and now we're only getting to point number 1 on the diagram, which is, well, different platforms have different intrinsic error rates. So I hope I've sort of put this in perspective. We're talking about quality metrics that change sensitivity by up to orders of magnitude, and we're not even actually doing the sequencing reaction yet.

Finally, we're going to talk this morning about sample types and validation. There are a number of different validation samples that you can use, patient samples, cell lines, and even engineered constructs. Leaving aside for the moment the local sequence context that can be changed that may introduce artifacts into these validation specimens, it raises the question, as Liz touched on, what is the

appropriate mix of mutations that you need in a validation specimen?

While it is true that next-generation sequencing can detect thousands, hundreds of thousands, millions of different mutations, there are not all in the same sample at the same time. So if you try to come up with a validation specimen that has dozens of different mutations in the same gene, in the same exon, at the same type, we run the risk of creating an artifactual validation scheme that doesn't really represent what's going on biologically. And that's important to keep in mind, especially when you start asking questions about, well, what about thresholds, varying allele frequencies? How do you now look for all these different mutations in these genes at different allele frequencies?

So at some level, it becomes possibly unsustainable as well as artifactual. And this is the nice segue, this is the last point I'll make, is where do you stop using biologic specimens to ask questions about assay validation, and do you move into in silico models? Once a laboratory has demonstrated that it can

make nucleic, extract nucleic acids, at what point, after the lab has demonstrated that there are no intrinsic biases to that process, can you say, okay, now you've demonstrated the wet lab point and now segue into what Panel 2 will talk about, is maybe using in silico data sets?

I know some of you are rolling your eyes about this. But it may be better actually to ask questions about varying allele frequencies and mutation mixtures by actually just sending labs in FASTQ files once you're sure that they can actually met nucleic acid. So with that, I'll quit. Sorry I ran over.

DR. SCHETTER: Thanks a lot. The next speaker will be Dr. Dara Aisner, assistant professor at the University of Colorado.

Presentation - Dara Aisner

DR. AISNER: Thank you very much for the opportunity to speak today. It's an honor to be here. I'm just going to speak also about some of the issues affecting pre-analytical processing, and I think about this as confronting and mitigating pre-analytical variability. So I'm not going to go through all of

this in detail because there is quite a bit of it here. At this point, I'll go through a little bit of it step-wise.

When we think about everything that we need to do to get to that step 1 that John just talked about, there are a lot of steps that get between a patient with a lesion to having DNA. And it starts with the tissue-acquiring procedure. It's modulated by the organ tissue, how it's handled in the laboratory, how it's processed, leading to how it's assessed, enriched, extracted, how that extract is assessed, and then how that extract and its assessment is used to modulate input into a library generation procedure.

So I'm going to go through each of these in a little bit of detail to try to understand how we think about confronting and mitigating these variabilities. In a lot of ways, all we can really do is confront them because in a lot of ways, there isn't that much we can do to mitigate them.

For example, in a tissue-acquiring procedure, you've got variabilities that arise from what its immediate transfer medium is, what's the temperature,

what's the cold ischemic time? And as John already showed us, we like to think that these things cause extraordinary variability, but I think a lot of data has shown us that for some things it does, and for other things it doesn't.

I think, as already been expressed, these are things that are not likely to be stipulated in a working medical environment. Really, truly, we have a very hard time doing this. And we've already seen examples of this. So for example, in the world of practicing pathology, we have worked very hard to establish parameters around the handling of breast samples for the subsequent ER, PR, and ERBB2 testing, and to some degree this has worked, and to some degree it hasn't.

The reality is that no matter how hard we try, there are always going to be exceptions that fall outside of these stipulations. So no matter how much we want to stipulate these things, we really can't. And as a pathologist who works in a anatomic pathology laboratory, I can tell you that just trying to apply these processes to the very small component of our

practice that is breast has created an unbelievable amount of turmoil and really having to turn systems upside down just to accommodate this one piece. And when we try to think about locking down these pre-analytical variables across the board, it at least at this moment seems to be outside of the scope of feasibility.

Here, I just show you some ER and PR, which has really been the goal of trying to standardize these things like cold ischemic time and formalin fixation time. It has led to some improvements, but even after these improvements, we do see variability. We do see things that fall outside those expected criteria.

One of the major mitigating factors from my perspective is that these sources of variability are less likely to impact DNA based testing than other testing. In my mind, the major solution here is to identify quality metrics that start after this point in the process, and that's going to be a message you'll hear from me a couple more times in this presentation.

Organ tissue. When we think about organ tissue, we think about, really, the matrix effect of

the organ. I want to pose a question here. To what degree is it necessary to validate tissue origin separately? Is lung really that different from skin, really that different from liver? Are these things really that different?

I'm just throwing up a picture of some lung and some skin here, and as somebody who looks at these for a lot of my life, I will tell you that I tend to think of these things existing as core constituents. Pretty much every sample that we look at, with some exceptions consists of core constituents, epithelium or parenchyma, whether it's neoplastic or not, inflammatory cells, red blood cells, and stromal cells.

From my perspective, lung and skin and liver are largely equivalent in how they are handled and what the ALK is. Really, I don't think we need to think about different organs as much as different matrices with the potential interfering impact.

So what do I mean by different matrices? Well, we've got tumors that like to make things. Right? Some tumors like to make mucin, and some tumors like to make chondroid. Some tumors like to make

melanin. Those to me are the things that are much more likely to have a matrix effect in an assay than what the origin of a tissue is. So just for the sake of fun, because I'm a pathologist, I had to include some pictures of chondroid and melanin.

In my opinion, validation should focus less on distributing across tissue types and more on matrix effects outside of core tissue components. I really think we can look at core tissue components as a singular entity.

Now, getting to this next area is an area of a lot of discussion. You have your fixative solutions; how long things are in fixative; when you process do you process using heat and pressure or do you process using microwave; how do you handle cytopathology specimens. This component of how cytopathology specimens are handled could take up three slides in and of itself in terms of the variability that's applied.

So the extent of variability here is very substantial, and I really believe that overprescribing these variables will lead to lack of access to testing for substantial proportions of patients. If we say

that samples have to fit inside specific boxes in order to be eligible for a test, what we're really doing is excluding a lot of samples. Also, if you try to overly prescribe these things, you'll interfere with the diagnostic practice because these samples are also overlapped into diagnostic practice.

So again, I'm going to come back to the mitigating factor here. The mitigating factor is that there are mechanisms to evaluate nucleic acid integrity. And I will say that again. To me, the solution here is to establish metrics that look at the resulting products, i.e., the nucleic acid integrity or the NGS data, or ideally a combination of both.

Coming on to specimen assessment, this is where we subjectively assess tumor cellularity. We subjectively assess the best approach for tumor enrichment. We subjectively assess the total quantity to utilize. This really is the practice of medicine. As a pathologist, this is how I take my training as a pathologist and integrate it with my understanding of the molecular assays.

Professionals who evaluate this and do this

for a living, I guarantee you do a much better job at this than any set of instructions you might find in a handbook. So just to try to make that a little more granular, this is actually from an FDA approved package insert for an FDA approved companion diagnostic.

Here, you can see that it has been stated out in three steps how to triage a sample for deciding how to handle it. I can tell you as somebody who does this for every day of my on-service life, if I tried to explain to somebody how to make these algorithmic decisions, it would take a lot more than three steps.

Now, for tumor enrichment, John already confronted this a little bit, but I think that we need to deal with this in some detail. Labs do use different approaches. Some labs don't use any tumor enrichment at all as was described. People just take scrolls off of slides. Some places will core or use a macroscopic isolation from the block. Some people use microdissection using glass slides as their guidance. Some people will microdissect under a microscope. And then there's a whole host of approaches for cytology specimens.

This is also the practice of medicine. Making these decisions about how to do this, about what is the right amount to get in there, about how to get the tissue from the block or on the slide, into the tube, really requires judgment and skill. So combining the visual assessment of the tissue, the means enrichment is a medical decision. And I would make the analogy that is much like a radiologist makes a medical decision for a patient based on what he or she sees.

I really, again, think this should not be overprescribed, and you really cannot assume that a sample can always be run without tumor enrichment or with only macrodissection. And I noted that in all of the FDA approved package inserts for companion diagnostics, macrodissection is prescribed, but I'd like to show you an example here.

This is a lymph node. It has metastatic melanoma in it, although it's not perfectly obvious at this power. It's not even perfectly obvious at this power either, but that's where that melanoma is. And this is what it looks like after we've gone in there with a scalpel and teased it out, and this is how

macrodissection does not cut the bill in some cases.

For this particular case, we microdissected. We actually used 12-consecutive micron levels. This is about a 45-minute to 1-hour procedure. This is what we do to make sure that the patient gets what they need out of the tests. This specimen would have had a very high probability of false negative without microdissection. Macrodissection really would have been insufficient, particularly when you think about the nuclear density of all these lymphocytes surrounding those tumor cells. We did identify a BRAF mutation in this patient with melanoma and went on to respond to vemurafenib.

In terms of extraction, you can look at different ways to extract: DNA only, RNA only, total nucleic acid. There are dozens of kits out there. This I think is a really key point. Laboratories need flexibility for the extraction approach. And while it's nice to think that we might want to say, oh, well, we know this specific extraction approach works for this kit, I think that the problem here is that it really ties things in a very negative way.

Inflexible approaches will lead to rapid tissue depletion. For example, if I need to use more than one kit procedure, and I have a small specimen, and each one needs a different extraction methodology, I have to pick between one or the other. There's not going to be enough tissue for both. It will leave me with an inability to perform orthogonal or backup assays.

So let's say the sample is insufficient for my next-generation sequencing assay, and I want to perform targeted backup assays for the patient. If this has been pre-prescribed in terms of the DNA extraction and my assay for, say, EGFR doesn't use the same DNA extraction, I'm locked out from using that.

Again, this will sound familiar. The mitigating factor here is that there are methods to evaluate nucleic acid integrity. And to me, the solution is to establish metrics that look at the resulting product -- that is the nucleic acid extract -- and/or the next-generation sequencing data. I do have a couple more slides. I apologize.

In terms of assessing that extract and making

that determination for input, particularly when we have limited tissue, you have one chance to get it right. So proper nucleic acid assessment is an integration of multiple pieces. It's a fluorometric quantification. I make that point because spectrophotometric really is insufficient for next-generation sequencing; microgel analysis, real-time PCR.

I will add that there is a subjective component here. So why we can define formulas that say "if then" and "if this amount with this quality, then do this," that works for a lot of the cases. But there are always exceptions. There are always samples where you go, hmmm, that doesn't quite fit the rules. I'm going to have to figure out how to try to make that work.

So there are examples of this, cases that never should have worked, but did; cases that clearly should have worked, but didn't. So the experienced practitioner can use these metrics as guidance, not gospel, for challenging cases.

To me, mitigating variability is how we go about implementing our practice. There are many

instances when a specimen does not fit into pre-established criteria. I believe we owe it to the patient to take a "try anyway" approach. So eliminating the ability to handle exceptions will eliminate access for many patients. Then of course the question is, if you are taking a "try anyway" approach, how do you really feel confident in the accuracy of those findings? I would argue to you that the key here is the ability to review the primary data. And being black boxed out to the data is a major detriment to being able to assure the quality.

In the setting of somatic conditions, a standard thing that we think about in germline, the transition to transversion ratio, really is not meaningful. By the time you have FFPE and the somatic condition, this ratio I think becomes much less meaningful. You can come up with a number of hard metrics that you can apply to a sample or to a run; for example, mapped percentages or off-target percentages.

This is a quality metric that we devise specifically for the assay that we're using. I don't have time to explain what displaced reads mean, except

to say that for this particular sample, this sample performed very well on that quality metrics. So we can establish hard metrics, but some of this is instinct, training, and clinical correlations, which is the practice of medicine. I'll make that point to you with a clinical example.

This is a 57-year-old female with adenocarcinoma of the lung. The sample was scant. The FFPE QC was poor to moderate, and the NGS result showed an extremely high level of artifact. How high a level of artifact? That's just page 1 of the artifacts. Each one of these lines represents a low level "mutation," which we were able to determine, based on our bidirectional design, that each one of these was actually an artifact.

Like I said, that's just page 1. You scroll through several pages of these and, oh, wait, look. There's something hiding in there. What's that you say? That's an EGFR exon 19 deletion.

Now, there are multiple places where this assay could have been cut off at the heels. Somebody could have said, "Oh, the sample's too small; don't

try." Somebody could have said, "Oh, that QC score doesn't work. You can't trust the results from that." Somebody could have said, "Oh, this doesn't meet the algorithm for the amount of noise in here. Throw out all the results."

There are multiple places where if you used an imposed metric, we wouldn't have been able to get an answer for this patient. This answer did turn out to confirm using a orthogonal assay. The patient's been on therapy. And the ability to test real patients in real situations relies on us using training expertise and judgment and the ability to review the data.

So in summary, the extent of pre-analytical variability is extremely high and cannot be regulated through mandate. Many of these are medical decisions. To me, the starting point for consideration of NGS oncology testing should be at the level of extracted nucleic acid.

I showed you a couple of metrics. I think that we have a lot of room where we can identify and devise additional metrics that can work as surrogates for stipulating extraction methodology or the preceding

steps. Assays that are black boxed to the data really increase the risk that subtle findings will not be identified. I apologize for going over. Thank you very much.

(Applause.)

Panel 1 Discussion and Questions

DR. SCHETTER: Thank you very much. That was two excellent talks that bring us into the panel questions. I think that both talks introduced a lot of the analytical variability that happens in clinical practice. We're going to shift to the -- the questions are going to be focusing on manufacturers should be required to address all of the different variabilities that are in clinical practice in order to show that their product could actually -- or at least describe how their assay performs under those conditions so laboratories are appropriately informed of the assay performance.

I have a series of four questions. We'll try and spend 10 to 12 minutes on each question, and then hopefully we still have some time at the end for audience questions. The first topic is on the quality

metrics.

Deviations and sample preparation and processing have large effects on the output of NGS-based assays. It is likely that manufacturers will develop NGS oncopanels for the use across the variety of sample processing methods and nucleic acid extraction protocols.

The FDA review process for traditional molecular assays is generally required that manufacturers demonstrate that an assay be robust across a variety of pre-analytical conditions that are expected for that assay. Therefore, the FDA review will consider evidence from the manufacturer that NGS oncopanels are robust across these conditions. The FDA is seeking panel input on the critical quality metrics that should be evaluated to ensure that NGS oncopanels will produce reliable results over manufacturer's stated range of pre-analytical sample processing conditions.

Based on your experience -- actually, to introduce the rest of the panel -- it's rather rude of me. In addition, we actually have Rajya Luthra, who's

director of the molecular diagnostics laboratory at MD Anderson Cancer Center. We have Michael Berger, associate director of the Marie-Josee and Henry Kravis Center for Molecular Oncology at Memorial Sloan Kettering. And we have Michael Rossi, who is assistant professor at the Emory School of Medicine.

So now to get to the specific questions, based on your experiences, what quality metrics do you think should be used to evaluate if nucleic acids are suitable for NGS assays prior to library construction? For example, nucleic acid concentration, nucleic acid purity and/or integrity of nucleic acids. How do you ensure that these metrics result in reliable variant calling?

For pre-analytical validation, what steps of the NGS workflow do you think should be evaluated?

What quality control methods from the sequencing run are most important to be evaluated, and how do you use these metrics to ensure the assay performed adequately? Do you think these quality control metrics are generalized in a manner that would allow the FDA to ask different manufacturers to provide

similar quality metrics? Under what circumstances do you think pre-analytical validation would not require evaluation of variant calling accuracy?

With that, I can open up the panel discussion, and I guess you can feel free to --

DR. PFEIFER: Well, since nobody is going, I'll start, just to play the role of provocateur. It's a very interesting question. I know labs do a lot of different things here, but in fact -- and this sort of mirrors what Dara said -- is we've taken a very "the proof is in the pudding" approach to this. You can spend a lot of time actually worried about the quality of the RNA and the DNA that you've extracted. And we basically measure the quantity of nucleic acid that's there.

Now again, the problem with that is, if we have a low input quantity when we do our library preparation, we will add some extra rounds of amplification in order to make enough DNA to do the assay, aware that that's going to change our library complexity, et cetera, et cetera. But really, the way we address the pre-analytical quality metrics for

nucleic acids is by making a library and sequencing it, and then looking at the metrics of the sequence results we get because there's a lot of information in those sequenced metrics that come back and tell you about the quality of your nucleic acids. It has to do with the quality scores of the individual base calls. It has to do with how they map.

So really, what we do is we go to the third one, the quality metrics of the sequencing run, to get some information about the quality of the preparation. And those are three things. Number one is we look at percentage of reads that mapped. So we know that if there's a problem with the library preparation, which can run amuck even if the nucleic acid is bad -- so it's the percentage of reads that map.

The second thing we look at in our assay is the percentage of reads that are on target. So there's a difference between the number of your reads that actually mapped to the human genome. The second one is the number that actually mapped to your target. We know that since we're doing enrichment, that the standard ratio.

The third thing that we use -- that, again, the reason we can do this is because we use a hybrid capture based assay, and we do paired sequencing. We can actually determine the percentage of unique sequence reads. And that turns out to be very powerful because that gives us an idea of library complexity. And again, it doesn't do much good to sequence the same genome over and over and over again from a tumor sample. What you want to do is actually look at a range of the cells within the sample.

So in our assay, usually -- we have a minimum metric for that percentage of unique sequence runs that are required in order to meet metric -- in order to go on and interpret the case. And it's a very important point. And I'm not going to argue that this is correct, and if people want to come up to me afterwards and point out ways in addition that we know that may be suboptimal -- but it's basically saying really it doesn't matter if you can make 50 nanograms of DNA or not, and really it doesn't matter if you can run it on a gel, and it looks like that you have high weight DNA.

What really at the end of the day is, you

want to make sure that the sequence that you got is indicative that it maps, it's good quality, that it's on target. And for us, that the complexity is there so that when we assign a variant frequency, we think it's representative of what's in the tumor rather than an artifact.

The problem with that -- and I'll just disclaim it so that everybody knows it -- is it doesn't address any of those things that I talked about to begin with. It doesn't talk about where did we sample the tumor or did we sample tumor are met. It's only for that area that we sampled, and we know that that is likely to be different elsewhere in the tumor.

It also doesn't address this idea about how we actually enriched the tumor. Right? It doesn't say anything about -- intrinsic in that, there's no information on the tumor cellularity, whether it's 95 percent tumor or only 20 percent tumor. You can't tell that at the level of DNA sequence reads. So it's a metric that you can use to talk about the quality of your assay, but it doesn't incorporate critical information about that case in terms of sensitivity.

DR. BERGER: I'll jump in if I can. I think John articulated everything that I've been thinking better than I can, so I'm glad he started. We do things almost exactly the same way. So we don't typically rely on pre-analytical QC to credential samples to move forward through the assay. We captured a lot of data up front, so our main panel that we run, our main oncopanel, is about a 400-gene assay that we've run on almost 10,000 samples.

So we've generated a lot of data, and up front we were generating a lot of this pre-analytical QC data. But when it became obvious to us that we weren't going to be using that as go/no-go decision, we stopped doing it. And we rely almost entirely on the post-sequencing QC, a number of those metrics that were described and some others that I can describe in more detail.

I think the original question was how can manufacturers establish that their assay is robust for each of the conditions that are being tested. I think there may be different levels of robustness, and that may be okay. There might be circumstances where the

assay can perform successfully 99 percent of the time and other conditions where it's 90 percent of the time, and others where it's 50 percent of the time.

In our lab, we still want to move forward even if there's a 50 percent chance of success as long as from the post-sequencing QC, we can clearly identify the success or failure or the limitations on our ability to detect certain mutations. Not all successes are even equal.

So given the level of coverage that we observe or able to generate, or the library complexity that we observe in the same ways that John described, we may recognize that when a given data set, our sensitivity may only go down to alleles of 15 percent or 10 percent allele fraction rather than 2 percent allele fraction. And that's okay as long as we can recognize that ourselves and report that back, because I think we don't want to deprive patients of the opportunity to receive a meaningful or clinically significant result simply because a sample didn't meet all of the strict thresholds that we would apply up front that would have prevented things from moving

forward.

I think that's the same theme that's already been articulated twice very well. I can go into more detail about additional QC metrics that we calculate, but maybe we can come back to that if other people want to speak.

DR. LUTHRA: Both John as well as Mike covered most of the things we do in our laboratory, too. When we strictly adhere to the concentration cutoff, DNA concentration cutoff, in our experience -- and will even be implemented, NGS in our lab -- we found that we would fail at about 10 percent of cases, so we wouldn't even go test them. But when we reduced the level of DNA concentration, now our statement is from 85 percent to 95 percent. So I think we are all talking about the same thing. We cannot have very strict guidelines that if it doesn't meet this concentration cutoff, do not proceed.

What I would like to say is I know we touched on several issues about microdissection and all that. In addition to that, I think an isolation technique includes some kind of cleanup, which is a magnetic bead

based cleanup or column based. And when you have, that would help getting a good quality. Of course, even with that, sometimes it's very degraded. There's still a possibility depending on how the tissue was fixed or processed, but we can reduce some of those things.

I think if you were asking our opinion of what should be advised, I would say if a company would like to double up their panel, they should look at several common isolation methods and compare them, and then say you could use any one of them. It's not restricted to only one procedure. I think that is important.

We also in the lab look at the library prep rate, the quantitation, and other parameters both John and Mike talked about. I think we also try -- as Dara was talking about, we want to make sure when the specimen is very limited and there is only one slide on a lung biopsy, are we going to say no because it's so tiny we're not going to test it, and we're not going to give a result?

So we try. That is our policy. And of course you have to make sure all the other parameters

are met, and you gave a very good example. That's what we have to look at the end product, from the beginning to the end. Thank you.

DR. AISNER: I'll put a slightly different spin on it in that we do have metrics in the lab that we try to work around. And as I sort of mentioned, we use them as guidance, not gospel. But there are cases where when we look at the yield and the quality scores on the specimen, we have a high degree of confidence that we have a strong suspicion that the library preparation will not work, and we know that we're looking at the very last of the material.

So in those situations, we are confronted with a decision of whether to go ahead and try the next-generation sequencing versus a flip to less complex targeted based testing on behalf of the patient because from a patient perspective, I'd rather get two specific answers than none.

So there is a subjective component, at least in our lab, where we do take a look at it at the upfront stage to decide, ooh, if we put this on next gen and it doesn't work, have we compromised the

ability to get any answer for this patient. And in many cases, what we can actually do is use the maximum volume needed for the next-generation sequencing and still have enough left over to try parallel approaches, which we actually do literally in parallel despite the cost consideration, so that way we're not impacting turnaround time issues.

For example, a patient with melanoma where the sample is marginal, we might say, okay, we're going to go ahead and try the next-generation sequencing, but in parallel we're going to do a targeted BRAF assay. So there are rare cases in our lab where we will just make the decision not to use the material on next gen because of our pretest probability thinking that it will not work, and we really do want to get an answer for the patient.

DR. ROSSI: I agree with everybody. I just want to add a couple of comments because I think this is very DNA-centric, and I think a lot of clinical laboratory testing is very DNA-centric. There's a move to do more RNA based assays. I disagree that that doesn't require any pre-analytic. I really think that

you need to do assessment of the quality of RNA if you're going to do an RNA based input.

I agree with this panel that it's very hard to establish definitive cutoffs for a particular assay based off of clinical parameters, but I definitely think that bioanalyzers should be used at least in that initial QC for RNA.

DR. BERGER: I agree. I agree with pretty much all the points that have been made. I want to emphasize one in particular. It's already been made, but I think it's critical. And that's the tests that use amplicon capture or PCR based capture and tests that use hybridization capture are fundamentally very different, and the metrics are very different.

Even the terminology is different. We talk about sequence coverage differently. For hybridization based capture, coverage often refers to the number of unique molecules, or template molecules, after removing duplicates from PCR, whereas amplicon capture, by definition everything is a PCR duplicate.

So in an amplicon test, you can say you sequenced a 10,000 X coverage, but that certainly

doesn't necessarily mean you sequenced 10,000 unique molecules. So your sensitivity may not be 1 in 10,000, or 1 in 1,000, or even 1 in 100.

I think that's one point. I wanted to come back to, as a second point, some of the specific QC metrics that we use. Part of it is based on additional content that we've engineered into our specific tests, but it's very important for us. And that relates to contamination.

So we aggressively monitor potential sources of contamination using a number of common SNPs that are scattered throughout the genome so if a patient is homozygous for those sites, we would expect zero percent frequency of the alternate alleles. But if we detect that for some reason, that may fail a sample, and certainly the lower input samples are more susceptible to this type of contamination. But we can use that to fail a sample. We can use that to assess the level of contamination that can modulate the level at which we're confident in the mutations that we call and that those are true positives rather than artifacts.

We've seen so many weird things. We see patients that have a prior bone marrow transplant and now are coming in with a solid tumor, and that's real biological contamination. We've had a patient with a tumor on their transplanted kidney, and that leads to strange artifacts. We've seen a number of things, and we've sort of seen them all. I think when viewed in the larger context, we can make sense of it and modify our calling criteria based on that. And we would have been otherwise blind to it I think if we hadn't engineered these extra steps.

So we sequence tumor normal pairs regularly, so we look for mix-ups between the tumor and the normal based on these SNPs as well. But I think more importantly, we use it to aggressively monitor sources of contamination. And that I think is going to be lab dependent and context dependent, and it's not all artifact.

DR. PFEIFER: I want to follow up on that because that's a very important point. In the larger context, this concept of contamination is called specimen provenance issues. And you can have specimen

mix-ups beforehand, and you can have contamination introduced into the assay.

There are a couple papers out there in the literature, and it's interesting -- you guys should write up the way you guys are doing this -- where people essentially look at the number of haplotypes. You're essentially doing a haplotype analysis.

This is a very important point because a few blocks from here at the NIH, most clinical trials require a specimen provenance step in there to make sure that you're actually testing the sample from the patient who's going to be enrolled in the trial.

It's interesting that many, most NGS labs don't rigorously actually have a provenance step in their assay. And what's intrinsic into what Michael just said is you don't necessarily need to add another step into your assay. There is bioinformatic ways to actually tease out the evidence of contamination in an assay.

So this is one of the things that I think is important. We also see evidence of patients who have bone marrow transplants, but we've caught a few samples

that actually had significant contamination in them because of this. So that's a good point.

DR. SCHETTER: Thank you. We're a little behind, but I just wanted to ask a quick follow-up on that. As manufacturers are going to be required to present -- or should present some metrics to show that their assay can work under the variety of conditions, do you think that it's possible to use some sort of coverage, like depth of coverage sorts of -- or what sort of metrics would be useful?

So if they change a variety of conditions, they're going to want to show that the assay works across those conditions. What would be the types of metrics that should be evaluated, and is it possible that it could be used for multiple different types of assays?

DR. BERGER: I think coverage is one of the most important, not just the total depth of coverage but the uniformity of coverage across all the target regions. But again, it's the terminology that may differ for hybridization capture and amplicon capture.

For hybridization capture, where you can

accurately assess the number of unique templates that you've sequenced, coverage probably is more important than anything, and that's what's going to set your detection sensitivity. But at the same time, you may have two samples, and one doesn't reach the same coverage as the other, but maybe that's because microdissection was used prior to the sequencing and that reduced the amount of input DNA but significantly enriched for the amount of tumor derived DNA.

So it may be okay. You may learn a lot from a sample that's only covered to 100X or 200X rather than 1,000X if the tumor parity is high enough. And I wouldn't want to use that necessarily to limit the ability to report results back, but I do think demonstrating adequate coverage on target, based on unique template molecule sequence, may be the most important thing I can think of.

DR. SCHETTER: Thanks a lot. We'll move on. I think both Drs. Aisner and Pfeifer touched on this quite a bit in their talks, but becomes a sample processing.

FDA is seeking panel input on the essential

pre-analytical variables that should be tested by manufacturers in order to claim their assay is sufficiently robust. With that in mind, what are the specific concerns you think should be addressed when evaluating how variation in sample processing for formalin-fixation and paraffin-embedded samples may affect the output of NGS oncopanels. I think that was already touched on.

What level of validation do you think is needed to support FFPE, fresh frozen and cytology specimen claims? How should differences in tumor cellularity be accounted for in pre-analytical quality control parameters?

If nucleic acid isolation methods are not specified, what evidence should be required to demonstrate any nucleic acid isolation method can be used?

DR. PFEIFER: Well, I tell you, we are extremely worried about this business about sample enrichment, and we worry about it a lot. And the reason we worry about it is for the reason that I pointed out in my talks, is it has a huge impact on

specimen -- or on assay sensitivity at two points.

Number one is just how many tumor cell nuclei are you putting into your assay? Is it 20 percent tumor or 100 percent tumor? Let's face it. There are those cases where it's only 20 percent cellular because there's a lot of stroma in the background. And even if you enrich for areas of tumor, you're still only at 20 percent tumor cellularity. Then there's this issue -- and Michael and I, we're mind-melding here this issue of library complexity. And we do sample -- we microdissections or coring so that we get a highly complex library.

Now, the reason I go into that is because it impacts in a sort of synergistic way the sensitivity of your assay. So I agree with Dara that there's a practice of medicine here, but on the other hand, even the practice of medicine has best standards, right, has standards of care.

So I agree with Dara that it's the practice of medicine, but from a guy who's involved in a lab that does this, we need to strongly consider where there are minimum thresholds because the concern is

that labs will take the same specimen and do different things, and come up with a different variant call.

One lab may test it and say it's below 5 percent. It's not there because of some threshold. Another lab may say it's there at 50 percent. So it can have a huge impact on patient care. Somehow we have to figure out a way that we can get some clarity, some reproducibility.

Just to show you how concern about this we are, after we mark the slides or we microdissect, those slides go back to a pathologist who then re-reviews the slides after the microdissection or coring has occurred. And that's part of our paperwork that goes with the case, is not an attestation but a statement that in fact that the area that was marked was collected, and so we're sure that we're actually sequencing this stuff that's there.

I know I'm a surgical pathologist. I know I have an intrinsic bias to worry about these sorts of things. But really, if you're trying to sequence tumor, everything that you can do to make sure that you're actually sequencing tumor increases the utility

of the test.

DR. AISNER: I agree completely. And although I do believe that those decisions are medical decisions, there are such thing as right and wrong medical decisions. I personally think that laboratories that are not enriching through specific tumor enrichment methodology probably are not making the right decisions for some of their cases. Certainly, we have seen through established literature that pathologists' estimation of tumor cellularity has variability within it. So I think that we owe it to our patients to overcorrect for that variability rather than undercorrect for that variability.

So I agree completely that we should be applying standards to those decisions. Just as John said, in our lab, we do the same exact thing. We retain all of the post-microdissection or post-coring material for a post-enrichment quality control. And there have been instances where I've sent cases back to start from scratch because I wasn't convinced that the right thing was put into the tube.

So I think that understanding that that piece

of the puzzle, as critical as it is, yet is very difficult to write down in a series of guidance documents do A, B, C, and D. It really is about creating the right guidance for the right audience.

DR. LUTHRA: I think I agree, a lab or any assay should have the quality metrics. They have to. And if there are deviations, it needs to be recorded, but we have to adhere to certain guidelines.

For example, in our laboratory, we say that if the tumor cellularity is below 20 percent, we do not subject them to microdissection. We would send it to laser capture microdissection. And also, it all depends on the sensitivity, analytical sensitivity of fewer downstream tests. So we do adhere to that, that we do not -- and then we would triage it to an orthogonal test rather than NGS. But as you mentioned previously, go to a single gene or whatever is indicated for that tumor type, we could at least salvage that way and give a report. I do totally agree that we do have to have the QC metrics for each step, and we should adhere to that.

I know it is very difficult because we are

dealing with specimens of different kinds which are coming from -- fixed differently and coming from everywhere in the states. So it's a very difficult question, but the guidance is something I think there should be a consensus on how to come with those parameters.

DR. ROSSI: I agree. I'm a PhD scientist, so I have to say that probably the most important thing is to make sure that a pathologist, a trained pathologist, is involved at the very start of this process. I think John and Dara did a very good job of trying to explain why that's the point.

If you've ever been on my end where you've sequenced something that wasn't tumor, it's very frustrating, and in a treatment situation, it can be lethal. So I think it's critical that the guidelines do stipulate that a pathologist, a trained pathologist, be involved in this process from the very beginning.

DR. LUTHRA: That's definitely -- upfront tissue qualification and training of the technologies for microdissection and re-review of the slides when in doubt. That has to be done under a pathologist's

guidance, yes.

DR. PFEIFER: One of the issues that I raised in my talk that I would be interested in knowing whether FDA is even thinking about is this question about where you're sampling the tumor. There are differences. We know there are differences within the tumor. We don't routinely do pair tumor normal testing; we just do tumor only. If you do -- and we have been asked, though, on several occasions by patients who were very knowledgeable, to actually sequence multiple places of their tumor because they want to know what the range is.

So now we start saying, well, okay, what exactly does that mean? Now, you're actually -- you know that you're getting a different result from different areas of the tumor, and we always do, and then which is the right one. So given that sequencing costs keep coming down and given that we're already being asked to do this, I'd be interested to know how often you guys have been asked to do this.

Now, we're moving into this situation where I want to know what FDA's thinking about what should we

doing -- how many times should we be sampling the tumor. Should we be sampling the tumor 10 times and mixing it all together in one tube. I'm not sure that's the right way to go. But we're going to be asked this question.

Then this question about the primary versus the metastasis. In our lab, we sequence the most advanced part in the tumor. So if a patient has a primary and now they come with metastasis, we sequence the metastasis because that's the thing. That's the cause of their disease now. But when they have metastasis at different sites, what is the site?

So when you have a set of guidelines about doing the testing, you can talk about -- if it's formalin fixed, you can talk about how much DNA you need, but we know intrinsically that there are differences in what we're going to find depending on where we're sampling.

So the interesting -- and I haven't seen it in any of the FDA labels that I've read that actually speak to that granularity. Is it intended for use only on a metastasis, or is it only intended for use on the

primary tumor, and that's an issue that we know is going to impact the results. So I know it's a little -- it's sort of a pre-analytical variable, if you will, but it's something that we know is going to change the result you get.

DR. SCHETTER: I can say that in the review of many different molecular assays, we are looking for tumor heterogeneity and we do ask sponsors to compare primary sites and metastatic sites. But the clinical claims from that generally will be coming from the clinical trial, and the language that surrounds the label is going to be based on what they see in the clinical trial. For example, if a clinical trial is run and no metastatic sites are used in the clinical trial, even if the analytical data shows concordance, you're still limited with what you can say about the metastatic site.

So I guess the question -- and in the scenario of what we're looking at now in which you're looking at multiple sites, multiple tumor types, multiple indications, the tumor heterogeneity issue is going to vary based on tumor -- not like tissue of

origin but each individual tumor. So it would be very difficult to address that.

I guess I would turn it back to you and what would be your recommendations to address that issue as far as what to ask the sponsors to do. Again, this could come back to a practice of medicine sort of scenario. But what sort of data would a laboratory feel comfortable with understanding that tumor heterogeneity that's present?

DR. AISNER: I would look at that from the other perspective. Rather than from the perspective of what should the submitting organization have to demonstrate up front, I would look at it from the perspective of in the constraints of what they submit, what will those downstream decisions impact in terms of the actual implementation of an assay.

If for convenience sake a submitting organization uses only primary tumors, does that mean that a laboratory cannot use metastatic tumors? So I think that, from my perspective, this is about how wording would be derived off of submission data so as to not overly restrict implementation.

DR. PFEIFER: That's an interesting -- so it's good on a panel to have people disagree instead of all seeing it -- so Dara and I are going to play a little bit of disagreement here.

That's an interesting perspective. I would be interested to know at least they're capturing that because it may be that there are intrinsic biologic differences between the primary and the metastases. We know that's true genetically. And if a manufacturer is looking for a label for a specific use of a drug, I think it would be interesting to know in what clinical setting has that drug actually shown utility.

If it was shown utility and people with stage 1 or stage 2 disease as part of like a primary therapy, it may have utility then, but by the time patients have developed metastatic disease, you could imagine the biologic scenario in which it no longer has utility. So showing that same mutation in a metastasis may no longer be aligned with the -- I totally loss my train of thought -- with the labeling or whatever.

So I think what's important is, I'm very reassured that you guys are capturing that because

that's going to turn out probably to be important in some settings.

DR. SCHETTER: Yes. And I think the most important thing is going to be to accurately represent what has been done with the assay in the label, so the labs can actually understand what's been done. And that way, they can interpret how to use that in their own way. But truth in labeling and accuracy in labeling is what we're going to be most concerned with. And we're not going -- we can't force everyone to do everything, but we need to at least accurately report what's been done so the laboratories can get an accurate assessment of what they're going to purchase.

DR. AISNER: John and I agree more than he thinks we do. But I agree that the biology is likely to be different in those settings. The issue from my perspective is that what is feasible for an organization to assemble for a submission is often different from the real day-to-day practice. And I would hate to see that a situation where what became convenient and feasible for the purposes of a submission then ties hands in the downstream

implementation.

So I think that there's a middle ground in here where one needs to understand the constraints of what is submitted without using those constraints to prevent implementation, I guess is what I'm trying to say. Ultimately for me, this is about realizing that what we're always looking at is the biology. And in my mind, the biology always overrides the test.

DR. SCHETTER: I think that's a very useful discussion. I think we're going to move on to the next one. And again, I think Dr. Aisner talked about this in depth in her talk and covered a lot of the points already.

As you know these NGS oncopanels, many of them are going to be coming in with pan-cancer claims, so we're seeking input on the essential pre-analytical -- should the expected NGS oncopanels be intended to evaluate multiple tumor types? Obtaining high quality DNA that is suitable for NGS assays is more difficult for some tumor types. And that's our assumption, but you guys can correct that.

Traditionally, FDA has required that

manufacturers validate their assay in each tumor type that is claimed in the intended use of the assay. Therefore, FDA is seeking panel input on the types of studies needed for manufacturers to claim their assay can be used across multiple cancer types.

Based on your experiences, what sort of representative tumor types do you recommend to be tested to justify pan-cancer claims or, as Dr. Aisner is talking about, what sort of cellular matrices should be tested?

What tumor types have been most difficult for you to get reliable NGS data from, and are there tumor types that should be excluded from pan-cancer claims unless the manufacturer specifically produces data from that tumor type?

Then, from a panel which is already existing, what level of validation should be needed to add or modify specimen types for an already approved NGS-based oncopanel, assuming that oncopanel wasn't already for pan-cancer?

DR. LUTHRA: In our experience with different tumor types, we had good reproducible results if the

specimens are -- resection specimens, you have a big tumor and you have good DNA. But when it is deep-seated organ biopsies, you have scanty tumor specimen. That's when we see that NGS doesn't work well.

I'm not too sure. Of course, if it is a bone specimen and we need a decalcification, that is a bigger issue. It depends on -- I think one of you touched on it, whether it is ethanol or methanol type of fixations and all that. Also, that if it's a strong assay versus a weak assay type of fixatives, decalcification method, those are all very critical when we are talking about bone specimen.

When we looked at several cases, though, we did -- our group has done a study and looked at the variety of tumor types. And we didn't find that there is any difference. It's not tumor type. It is what specimen, what DNA. That dictates and how it is fixed. That is more critical than tumor type. That is our experience.

DR. BERGER: And ours has been exactly the same. I think a much bigger determinant are the two

things that Raja just mentioned, whether it's FFPE versus frozen and whether it's a small biopsy or finding the last cytology specimen versus a resection.

Certain tumor types may have slightly variable failure rates because they more often come to the lab in one of those forms, not because of anything innate about the disease. Maybe a possible exception, we struggled in some instances with some prostate cancer. And the pathologist to my right can probably explain that much better than I can.

But our experience has been that the tumor types perform generally comparably well. And another point is, in our experience, we've sequenced over 60 different general types of cancer and over 300 very specific types of cancer. And it just wouldn't be feasible to validate each one on its own. So there are practical implications to this particular question.

DR. ROSSI: Yes. For me, I think the pan-cancer claims have to be established whether or not the pan-cancer panel is going to be run independently of other ancillary tests. I feel like you only have so much space that you can fit on the panel to get

adequate depth and quality of sequencing.

Many people would split the hematological malignancies from the solid tumors. That's because many translocations are very well known in hematological malignancies that may not be covered by a DNA based panel set up as a pan-cancer panel. So I think it's very critical, as Dara said. So I emphasize over and over again that we use genomics to get at biology. And you need to understand specific mechanisms that drive specific disease entities. So I think that's really important.

DR. AISNER: So I already explained in my talk my view on interfering matrices as opposed to tumor types. I may be stepping on the toes of the analytic piece here, but I really think that rather than focusing on tumor types we did, 10 lungs and 10 skins and 10 colons, and list them out, I think we should be looking at interfering matrices or the potential for various matrices to interfere.

But I think that, really, validation should focus on mutation types and locations. And I'm sure this will be discussed in the next panel, but I think

that when we think about how to justify a pan-cancer claim, we should be looking at the mutations that are most commonly seen in a wide spectrum of cancers and the fidelity of an assay to detect those.

I really do believe that when you think about tissue in terms of its core constituents -- and I think that this has been echoed by others on this panel, that our universal experience is that it doesn't matter if the tissue derived from lung versus colon, barring any interfering matrix effect, that it's really about the ability to detect the alteration, not about the tissue that it came from.

DR. PFEIFER: Right. I agree with everything people said. The only tissue type that we reproducibly or consistently have trouble with is specimens that have bone in them. And that's just because people use very powerful low pH fixatives or low pH to do the decalcification. You can do it with types of fixatives or processes that don't really impact nucleic acid such as EDTA, which we encourage people to do. But when cases come to us in consultation, we don't have any control over that.

I want to focus on something that Michael said. It's real interesting. What representative tumor types do you recommend be tested to justify a pan-cancer claim? And you put a very interesting spin on that, which is exactly what is your assay designed to do?

One of my pet peeves in this space is people will do a test and not actually call out in the test the types of mutations that their assay can actually find. So some of us who do hybrid captured based tests and spend a lot of time developing capture probes to find translocations within the same assay, as well as small SNVs or small indels and have validated it for larger indels, and then some rearrangements -- it's easier with RNA, for sure, than it is with DNA, but you can find some of the recurring ones.

So we do a very complicated assay that includes all these variant types. And then some other lab offers a test for less money that is nowhere near as comprehensive. And it is absolutely unclear to the ordering physician the differences in the range of the mutation types that that test is designed to detect.

I think that's a very important point in all of this. We're talking about all these pre-analytical variables, but there needs to be some indication somewhere as to the range of variance that this test is intended to detect because the bioinformatics to find large indels are different than they are to find small indels. And they can be very important in some tumor types. And yet, if you're looking at certain kinds of AML, if you haven't validated the test, clinicians may be unaware -- the people who order this test may be unaware that this lab test can't find those.

So that's a very good point, Michael. It's amazing that it hasn't come up earlier today.

DR. SCHETTER: So that topic is extremely important, and it's going to come up a little bit later. But we agree that -- I think the tests, we expect to have labeling and such that those sort of limitations should be present. So those discussions will be happening a little bit later.

DR. AISNER: I'll reiterate that bony specimens really are the major challenge. And we and many other organizations go to great lengths to try to

avoid decalcification. So when we know there's a sample coming in from a patient where it's a bony lesion, we actually have special handling procedures that work to avoid decalcification to every extent possible. We ask for paired FNAs because those will not have decalcification.

So there are pre-analytical steps that can be employed to try to increase the downstream quality of what you have with that upfront knowledge. But the real challenge there is in the communication streams that everybody, from the person who's ordering the test, to the person who's doing the biopsy, to the person who's receiving it, knows exactly how to handle it at every step in the process.

The other sample type that we do have difficulty with are highly necrotic samples, and I think that we get very noticeably increased background from our highly necrotic samples. And again, this is where microdissection comes in and becomes highly critical because scooping up a lot of necrosis can really alter your ability to sift the signal from the noise.

DR. ROSSI: If I can make just one more point. The other thing about enrichment -- this is a very solid tumor-centric panel, but -- so something like multiple myeloma where you can have a fairly low disease burden in the bone marrow enriching specifically, establishing a part of your assay, your workload, to enrich for the myeloma cells is critical because that will dramatically affect what your results are.

So I think that this does require input from pathologists and in the case of hematological malignancies specifically from hematopathologists. And I think on the surgical pathology side, it's a little bit more straightforward in terms of assessing what tumor percentage is. But for pan-cancer claims, I think it's very, very critical that we have this established sense of this is the disease, this is what the requirement is of cellular input, and this is how you then proceed with the assay.

DR. SCHETTER: Great. Thanks a lot.

So we're running behind. This next question will limit to about five minutes, so we still have a

few minutes at the end for public questions. It changes gears a little bit.

It's impossible to acquire clinical samples for all the possible variant types that may be incorporated into an NGS-based oncology panel, therefore many manufacturers are proposing to use contrived specimens to supplement their analytical validation studies. FDA is seeking panel input on how contrived samples may be used to demonstrate analytical validity of an NGS-based oncopanel.

What types of commutability studies should be conducted in order to infer the performance of an assay on clinical samples from data obtained from cell lines or plasmids, such as what quality metrics could be used to show similarities and differences between them? Would you expect to make calls with more confidence in contrived samples, and how could studies be adjusted to more closely mimic clinical scenarios?

When clinical samples and cell lines with specific variant types are not attainable and with the understanding that plasmids like the 3D architecture of genomic DNA, should engineered cell lines be a

preferred method of contriving samples for analytical validation purposes?

DR. PFEIFER: I touched on this briefly in my comments. The problem with all of these engineered samples, whether they're plasmids, whether they're cell lines that have engineered mutations, is if you look for the associated changes in architecture or sequences, you can actually back your way in to finding a lot of these variants.

Since we look at -- one of the problems with using cell lines, people say, well, engineer cell lines, and then you can mix them to look at different allele ratios and stuff. Well, if you use the same cell line for everything, you can avoid the problem. But oftentimes if you just mix cell lines -- if you're looking at SNV frequencies, you can be aware that there is a mixture of cell lines, and you can actually see what the ratio is. And then you can ask your bioinformatic pipeline to go find things that have that ratio. So you can validate an assay in ways that are completely unrelated to the biology of what you would do in routine clinical practice.

My view in this, as I said in my talk, is what's important here is to demonstrate that laboratories can make -- I'm not minimizing this. But laboratories need to be able to demonstrate that they can make, that they can extract nucleic acids of high quality. And that's the important point here. I mean, that's the proof in the pudding.

There is no doubt that there's a role for giving them wet lab samples to show that they can extract nucleic acids, make a library, and put it in their pipeline and find the variants at specific and hit specific targets. There's no doubt that that's true and that's necessary. The question is, is how is it sustainable to look for all these variants? Where do you draw the line for variants, different allele frequencies, so that you don't end up doing something that just is crazy?

The question to me is where you draw the line, per se, and can a lab do this wet lab piece? Can a lab make a nucleic acid preparation and get the right answer? Or maybe there are -- and this, again, segues into the next panel. Are there bioinformatic ways to

actually test your assay sensitivity when you actually get -- from the sequence information? So it doesn't do any good if you make the right library prep and sequence if your bioinformatic pipeline isn't appropriately tuned.

DR. LUTHRA: I think as we cannot have too much samples, we have to go to this approach of having the cell lines there are plasmids. However, we have to make sure that we mimic the real-life tumor sample material. So your cell lines, then you want to make sure that you fix them and number them.

Those are some of the important things we have to make sure we follow because you cannot have some cell lines which are fresh frozen, and then you are comparing your data with the FFPE samples. I know you're saying the lab is very -- I mean, the panel is solid tumor-centric. But we do a lot of hematological malignancies, and also you do have to have the cell line. Yes, we you can get a high molecular rate DNA and a good quality. However, there are situations, even for hematological malignancies, where you would have fixed tissue, so we have to worry about that. And

we want to make sure what type of controls we are doing.

In addition to validating with these plasmids or cell lines, I think we have to do a large number of patient samples in validation. That's very critical to have that before you go on to implementing the tests.

DR. BERGER: I agree that that's important I think. I don't want to diminish the importance of large numbers of patient samples, and that's the way we've done our own validations in the past. I do really like the idea of using in silico validation to supplement a laboratory validation. I think if we can do a reasonable number of experiments that hit all of the QC metrics that we're trying to hit, performance metrics that we're trying to hit, we've established the laboratory processes and the performance characteristics of the assay.

We can't possibly sequence every sample with every possible mutation and every context and every gene that we want. The only way to demonstrate that, if that really needs to be the bar, is through supplementing the laboratory validation with in silico

data, maybe control data sets that people use to validate their bioinformatics pipelines, maybe in silico mixtures of samples to supplement the experimental mixtures of samples that we are doing in our laboratory validations.

I would also make the point -- I don't have an answer to this -- of thinking about validating the performance for copy number alterations. And rearrangements can be very different from point mutations. Especially, these mixing experiments that we've used to assess the limits of detection for mutation calling are actually very problematic when you try to use it to assess limits of detection for copy number alterations.

DR. LUTHRA: That's actually a very good point we didn't discuss here because it depends whether you are looking at the copy number of you're looking at a single nucleic degradation, LOH. You were cut off for tumor -- a person's age will vary depending on the type of mutation you are looking at. That is a very important point we didn't discuss.

DR. SCHETTER: With that, thanks. We are out

of time, but I'll open it up for public questions to maybe one or two questions, if anyone has questions.

MALE AUDIENCE MEMBER: So there is sort of the practice of medicine, but there's also best practices in medicine. And I think if you look at the companion diagnostics that use nucleic acids, there's a pretty much uniform sample processing step that's in there in order to analyze a sample properly. And part of that is very likely due just to make sure that the assay in the hands of the end user performs according to the claim performance characteristics of that assay.

With that in mind, could a lot of these input requirements just be solved by educating the pathologist as to how the sample has to be processed?

DR. AISNER: Theoretically.

(Laughter.)

DR. AISNER: I say that practically in jest. But truly, in reality, the practice of pathology really has diverse conditions. And I probably about once a month get somebody who asks me, "Well, why won't pathology just move away from formalin anyway?"

The reality is that these are practices that

have been set in medicine for a long time. The example I showed of an FDA handbook actually also stipulated that the specimen should be fixed for between 14 to 24 hours. Now, the reality is that labs when they get a sample, unless it's a breast cancer sample and it's been stipulated that you record formalin fixation time, we don't actually know how long a sample's been exposed to formalin.

I think that if we want to talk about standardizing pathology practice at that level, it's a completely different conversation. Just the process of doing this for breast cancer specimens has really involved an enormous amount of effort on the part of pathology labs and hugely increased budgets for personnel. So the constraints that pathology labs work in right now, I would argue that it's really not feasible to record the ischemic time, the formalin time, the processing time, the processing conditions of every sample that is processed in the United States.

DR. PFEIFER: Well, I have a slightly different take on that perspective. At the end of the day, what we're trying to do is do testing that is

reproducible between laboratories and that is accurate. It's not just reproducible between laboratory, it's accurate. If in fact pathologists need to monitor the formalin fixation at the time that things are in formalin, then we need to do that.

If that has increased costs -- not if, it will have increased costs -- then this is something that we as pathologists will have to build into the cost of these assays. And now, we've just wandered into the whole payment thing, but there's a reality there. As we pay attention to more details, costs go up, and we need to recognize that.

What I tried to say in my presentation -- and I think I've tried to say a couple times now -- is we need to be careful that we're not focusing on something we're sort of locking the next-generation sequencing barn door after the next-generation sequencing horse has already been stolen. A concern about 80 percent tumor cellularity or only 24 hours in formalin is misplaced if someone can have 5 nanograms of DNA and just do extra cycles of amplification, and end up with a library that's low complexity, and we get a result

that has little clinical utility.

So I agree with what the questioner said, and I agree with Dara that there will be increased costs, but we have to stop talking about formalin fixation as the evil in all of this. We know there are intrinsic biologic heterogeneity. We know there are differences in how labs may or may not choose even if you passed a certain threshold to do an extra microdissection. And we know there are differences in the way that laboratories are preparing their libraries.

So I guess what I'm arguing for is -- what that question to me sort of makes me grind my teeth a little bit is it's focused too narrowly. Addressing formalin fixation will not fix the variability between laboratories and is a small component of the reasons underlying -- it's only one of many reasons, and a small reason, underlying the differences between the results the next-generation sequencing laboratories get.

DR. SCHETTER: Last question? We're already in the break time, so I think it's the last question.

MALE AUDIENCE MEMBER: First, I commend the

emphasis on DNA integrity as a mitigation for many of these variables. Thank you. On the topic of commutability between cell lines and the plasmids, you mentioned testing clinical samples. Do they have to have a variant, and how many do you usually test to show integrity of your assay?

DR. PFEIFER: We tested in validating our assay several years ago -- we tested over a hundred patient samples. But they were our patient samples. We shared some of them with other laboratories. But yes, it was over a hundred. I forget the exact number. And every time we introduced an updated assay or change our assay, it's dozens that we test.

DR. LUTHRA: Yes, we did similarly, but the main thing is to have a mix of different types of variants. Like if you're looking for copy number changes, you should have enough to do that and then translocation. So you do want to have all the -- but you cannot really every -- again.

DR. BERGER: Yes, the same for us. We sequence several hundred samples, cover the different classes of mutations and the different -- many

replicates -- sorry; separate samples with mutations in the genes that we were interested in validating for the purpose of our assay.

We are in New York State. We were adhering to New York State Department of Health guidelines, which was very instructive in this case, a potential model for future regulation because we were demonstrating on at least 10 positive control samples for each class of variant in each region that we were intending to validate.

DR. AISNER: And to get to the question of whether every sample you touch should have alterations, it always depends on what sample source you start with. So for our first round of next-generation sequencing validation, we had a deep bed of samples that had been previously tested that were negative for things we'd happened to test for.

Of course, in that process, we identified things in those samples that had not been previously known because they haven't been tested for, so we went and orthogonally confirmed every single one of them, and that's an iterative process. So our next round of

validation, our number of samples that have nothing in them is a lot smaller. We will of course use some of those as part of a proof of principle that by expanding the content of our assay, we are in fact picking up what we intend.

So you can use a biased -- by using samples that have not shown any alteration through previous testing, you can bias your knowledge base to understand that your additional content has technical value.

DR. SCHETTER: With that, thanks a lot. Again, thanks again for all of your input. It's been a very useful discussion. And with that, we're going to enter break. Thanks again. Be back at 11.

(Applause.)

(Whereupon, at 10:30 a.m., a recess was taken.)

Panel 2 - Donna Roscoe

DR. ROSCOE: My name is Donna Roscoe. I'm one of two branch chiefs in the molecular genetics and pathology branch. The other branch chief is Eunice Lee. She's here with us today. I'm very excited to be moderating this next panel on analytical validation,

and it's a really critical panel, that we have this discussion. And I'd actually like to start off by having the panelists introduce themselves, talk about their affiliation, and a little bit, just like one or two sentences, about your expertise with next-generation sequencing.

DR. HEGDE: Good morning, everyone. My name is Madhuri Hegde. I'm from Emory University. I am faculty in the Department of Human Genetics and Pediatrics, and I run Emory genetics lab. I am the executive director of Emory genetics lab. My experience mainly comes from inherited diseases and then spans over to the somatic area as well. I've been doing diagnostic -- involved in diagnostic analysis for a very long time, probably 15, 20 years.

At Emory, we were one of the first labs to get into next-generation sequencing in clinical diagnostics. So that's been quite exciting but have tried everything from all the different platforms that have been available. Thank you.

DR. VAN ALLEN: Hi. My name is Eli Van Allen, and I'm a medical oncologist at Dana Farber and

a computational biologist and cancer genomics researcher both at Dana Farber and the Broad Institute of MIT Harvard. I have an independent research lab focused on translating complex large-scale genomics into the clinic, and much of our lab is focused on computational approaches to interpret the cancer genome at the point of care for patients.

DR. DEIGNAN: Hi. I'm Joshua Deignan. I'm associate director of the UCLA molecular diagnostics laboratories. I've been in this field for about seven years now. And actually, our lab got our first foray into next-gen sequencing with clinical exome sequencing a number of years ago, and then very shortly after transitioned into somatic NGS, and now hematologic NGS as well.

So I think what I hope to accomplish today as part of this panel will be to compare and contrast a little bit what we've learned on the germline side with what we're now trying to do on the somatic side because I think there are a lot of parallels, and there certainly are a lot of differences as well.

DR. EBERHARD: Hi. I'm David Eberhard. I'm

a pathologist at the University of North Carolina. For the past five years, I built, grew, and directed the gPATH preclinical genomic translational pathology laboratory. Our laboratory is focused on providing next-gen sequencing based on characterization of clinical oncology patient samples to support clinical research protocols in the Lineberger Cancer Center at UNC, as well as for external clients and collaborators.

DR. KLEES: Hi. I'm Robert Klees. I work for the New York State Department of Health in the clinical lab evaluation program, and I'm the primary, pretty much sole reviewer most times, of all oncology assays that seek New York State approval. And as we've seen in recent years, there's been an uptick in the NGS assay, so I have a strong interest in seeing where this goes.

DR. ROSCOE: All right. Great. Thank you.

At this time, I would like to say that this workshop panel will follow the same format as before. We'll have two talks, 15 minutes for each of three sets of questions, and then followed by questions from the audience, 15 minutes for that. At this time, I'd like

to invite Madhuri Hegde up.

Presentation - Madhuri Hegde

DR. HEGDE: Good morning, everyone, and I want to thank the FDA for asking me to speak here today. We are going to be talking about the analytical side of next-generation sequencing as in really the part which the manufacturers are going to be putting together and the labs are going to take it in their labs and use it as a clinical assay.

Just to start with that, I'm going to talk a little bit about -- Josh just mentioned that the germline testing and the somatic assays are thought to be very different kinds of assays where we can draw a lot of experience from just what we have done in next-generation sequencing for germline testing and use that in developing the somatic assays. One of the areas in germline testing which we look for is mosaicism, so we are looking for very low allele frequency in a lot of different samples for germline testing as well. That kind of translates over to when you're doing oncology type testing.

This figure really shows you the differences

between disease diagnosis versus clinical utility in cancer diagnosis for next-generation sequencing. Here, they are really talking about a select set of mutations which are clinically actionable, that we want to develop some analytical parameters for it to be used in a clinical setting.

Again, to just point out the differences between cancer and rare disease, where once you have identified a mutation in an inherited setting, you actually have the mutation to go after, whereas in the cancer setting, you are probably doing some repeat testing. There's a lot going on, which is different. And we have just talked about the pre-analytical part of it, which is a sample type, which you probably are going to go back to the same patient and get a different sample from the patient to be tested again.

Let's look at NGS in oncology, the significant advantages we have in using next-generation sequencing over the traditional methods. There are many, many publications now which focus on that. But the real advantage comes down to the ability to sequence a large number of genes, or hot spots, in a

very high throughput fashion for many patients, allowing the simultaneous detection of single nucleotide variation, copy number changes, deletions, and duplications in one single assay.

This really comes down to the assay design itself in the analytical setting of how you want to design your assay and is absolutely critical. That's something that the manufacturers will focus on as this assay is put together. The other thing that is coming up very quickly is the application of known or approved drugs to new cancer types and number of different indications and new target discovery.

This is something I think the FDA also has to look at because the field is moving really fast, and cancer is a very time sensitive type of setting where the assay is being performed in. So this is something that has to be looked at because how are we going to include new targets in an FDA-approved kit that now needs some post-approval modifications?

Starting off with the different capture methods, there are now many methods that have been described in the literature. Probably two which stand

out is the hybridization based methods and the amplification based methods. We heard from the pre-analytical panel that these methods have some intrinsic differences by itself in how you design the assay. It also starts with the starting material where your DNA, are you just going to use extracted DNA or do some amplification procedures before putting it into the assay itself.

The design aspect of the assay itself is the overall design and the by-gene design. This was also touched upon earlier that the gene, the mutation, and the sequence context around the mutation, the flanking sequences, were important because you are going to determine the qualitative and the quantitative efficiency of your assay itself.

I'm not going to get into the fresh-frozen and FFPE specimens. We have discussed a lot about it. But it comes down to the integrity of the sample that you're going to put into your assay and the quantity recommended that should go into the assay, which is something that needs to be looked at very closely.

Now, there are some differences in the

approaches to detection of singular nucleotide variation and the detection of indels and other mutation types. I'm including translocations here because the design of the assay is absolutely critical here in terms of how you're going to detect a translocation, do you know the precise breakpoint, and how the assay actually has been designed, and what information is actually given on the label to the user of how this design has been put together because there are always exceptions and variations that happen in different sample types.

I'm not going to get a lot into bioinformatics because the next speaker is going to talk about it. So I'll sort of leave it there because the detection of copy number variation or translocation has a lot to do with the bioinformatic approaches as well. But a singular nucleotide variation is probably a little bit more easier when you look at the informatic approaches.

Again, going back to the advantages of detecting mutations in next-generation sequencing, you have an extremely throughput assay here because you can

look at many targets at once, and you can develop an assay which is unbiased towards the different types of mutations.

But again, going back to the type of method you're using, whether it's a hybridization based assay or an amplification based assay, will create that difference in your ability to look at the sequence both in the qualitative and the quantitative assessment of the sequence that is getting generated.

Now, the ease of this kind of technology is a digital readout. For those of us who have done Sanger sequencing for a very long time, we've been looking at chromatograms. Here, you're actually looking at a nucleotide and assessing the true quality of that assay itself, whether the nucleotide is present or not. This can be different when the sample is heterogeneous, so you have to develop approaches which are consistent across the board when putting something like this together.

Now, touching a little bit on the heterogeneous nature of the tissue itself, that has been discussed in the pre-analytical panel as well.

But the evidence based selection of the target is absolutely critical, and there is literature around it. This is really going to now the practice of medicine of what we include in the assay is evidence based, and there has to be solid evidence of why that target actually got into the final design of the assay itself.

The clinically actionable cancer mutations and the detection of those mutations at a very low allele frequency is probably the most sensitive part of this entire design. And how do you establish the validation of this type of an assay when you're going down to 10 percent, 5 percent, and the claims that are being made around it.

This is one question which comes up again and again, how many samples to use for validation before you can actually put something together and apply for approval? There are these numbers that float around. If you look at the literature, 20. We've already heard from our panelists earlier that some of have done 40, some have done 100, some have done 200.

When I look at how many samples have to be used for validation, it's really critical going back to

the design, the sequence, context, the mutations that are included in the panel, and that the validation should be around the type of assay, and it should be representative of all the mutations that have been included in the panel itself; multiple times, run inter-personnel -- there are differences that happen, especially when you're trying to detect a very low allele frequency. The idea is trying to sum up the entire analytical validation in two or three sentences. This is absolutely critical. Reproducibility studies are important, and the robustness of the assay is also critical.

One of the things that were discussed earlier is in silico approaches that can be used because it's really difficult to find samples validate everything that you're including in your assay. John has done a lot of work on this, how can you use different in silico approaches. But going back to the starting material and validating on that is also absolutely critical before you can say that this can be used across the spectrum.

There are papers which talk about use of

HapMap samples. I put this out there on the slide because these are samples that are easily available, and labs can use them when they're trying to run controls in their lab along with the assay that is already approved.

Deep sequencing is something we have heard again and again and again. In the literature, it is talked about. We heard this earlier as well. When I look at deep sequencing and coverage, I think from the manufacturer's perspective, it's important to determine the cutoff because if you have 10 mutations, 20 mutations, or 40 mutations, you want to be able to show that every mutation is going to get detected.

For some mutations, you might get some extra deep coverage to get that other mutation which is not getting covered properly. The reason I say that is the sequence context, the GC content might affect the ability to detect that particular mutation, which is heavy on GC content. So the flanking sequence is also important when designing the assay itself.

One of the questions that comes up is how do you compare this to the germline assays because when

you start getting an allele frequency of 50/50 between the mutant and the wild type, you are starting to go towards a germline evidence for that particular mutation. And how do you actually put that in your assay and determine the sensitivity and the specificity of that particular sample for that assay itself? It's absolutely critical that the testing is done or the validation is done across the spectrum when seeing that this assay is working at this depth, at this cutoff, when you are putting this out there.

I'm not going to talk about the types of specimens; we've already discussed that. But I think some sort of a deviation might have to be considered simply because getting a boxed [indiscernible] assay and running it in the lab is one thing, but having the specimen types and being able to offer a wide range of cancers, and then doing the post-analytical analysis and putting a clinical report together are different things from just using an approved assay.

Touching a little bit on the orthogonal methods can be used for confirmation, I think we all know this today, that if you detect something on Sanger

or NGS, you probably may not be able to confirm it by Sanger sequencing. And we have to kind of bring this up again and again because there are different platforms you can use for confirmation of the mutation you have detected in the approved assay, but Sanger sequencing may not be the right assay to use. Now, that increases your cost of confirmation, but that's something that has to be remembered as these panels are put together.

Again, going back to the optimal coverage depth through the validation of the assay and comparing that to the orthogonal confirmation method that is being used is critical when the validation is being actually performed.

This is my last slide. I think when I talk of limitations, this can also be looked at a little bit as advantages because I don't want to put big holes in what I've already said. One of the important things, I think putting it out there that what is the limit of detection is absolutely critical, the sensitivity and the specificity. In this case, there are many publications now that talk about going down as low as

10 percent. But if the assay cannot detect anything lower than that, I think that has to be disclosed up front so that there is no confusion over it.

One of the things which NGS allows us to do is to do multiple samples at the same time. There is this confusion of am I contaminating something by doing something like that or is this an ok thing to do. That has to be addressed within the validation process itself, the analytical validation of the assay.

There are many publications which have shown that the crossover in doing a library preparation is not a huge problem, but there could be measures that can be put in place to address that. And that also got sort of discussed a little bit in the pre-analytical session.

That was my last slide. I'm going to stop there and hand it over to the next speaker.

DR. ROSCOE: All right. Thank you for that excellent talk. At this time, I would like to invite Dr. Eli Van Allen up to the podium.

Presentation - Eliezer Van Allen

DR. VAN ALLEN: Great. Thank you again for

inviting me and giving me this opportunity to present. My name is Eli. I'm a medical oncologist at Dana Farber and a computational biologist. So I'm on the research side really trying to drive the implementation of next-gen sequencing in the clinic and also on the clinical side a downstream consumer of what we produce. So it's been interesting to see this evolve over the last two to three years in this space.

I'm going to talk for maybe about 10 minutes or so on a few different bioinformatics considerations that we're hitting as we're trying to implement this in the clinic and make next-gen sequencing possible for our cancer patients.

Just as a brief outline, I'll talk about some of the validation considerations for variant types and methods considerations looking at some of the tumor-only panel challenges that we're now facing as we consider tumor-only sequencing and matched sequencing, and then talk just for a few minutes about inferring global genome properties from next-gen panels because that's becoming increasingly relevant in the world of immuno-oncology.

First for validation considerations, and for that I'm specifically talking about how do you know whether the somatic alterations you're looking at are actually there or real. Depending on the assay that you have, as was mentioned in the first talk in this session, the opportunities to detect things range from simple hot-spot point mutations all the way to short indels, copy number alterations, and oftentimes for instance if you're using RNA based panels, looking at fusion products.

The level of analytical validation is variable for different components of this analytical pipeline. To really emphasize this point, I'm going to contrast somatic mutation with fusion detection. In the world of point mutation calling, there's been quite a significant amount of effort done at multiple institutions to really robustly create somatic mutation callers for next-gen sequencing panels.

This is the version that we use at the Broad called MuTect, which was developed by Gaddy Getz and Kristian Cibulskis. This is just one example, and I think there's plenty of other fantastic ones that are

out there in the community. They all sort of circle around the same kind of properties, where you can actually just demonstrate that the power to detect a given mutation is a function really of the sequencing depth allelic fraction of the mutation when you actually apply some of these modern mutation analysis tools.

In this particular study, they demonstrate that, and these being the different allelic fractions and what one is able to detect at different depths of sequencing. They also actually went through orthogonal validation for a set of different cancer next-gen sequencing studies to actually use orthogonal technologies to really highlight the validation rate for this approach.

Like I mentioned, there are other algorithms that are doing similar things with similar results, and I would actually like to highlight this effort, which is at the bottom here, where they are now in the somatic point mutation space, sort of crowdsourcing team efforts to try to actually figure out what's the best approach and the optimal approach.

Actually, this is a publication that came out of the TCGA ICGC collaboration, where you actually had I think over 10 teams around the world competing to actually come up with the right approach. The take-away I had from reading this particular paper was that there are some differences, but for the most part, people seem to be pretty confident in calling point mutations.

There are issues related to sample quality that are upstream of this, and of course I think documentation of how you're actually using these algorithms, which don't exist in a vacuum because they rely on panels of normals and other features that you may need to build into that algorithm that could change; so documenting that as key. But in a lot of ways, this is a pretty robust thing that we can do in the clinic.

I really point that out to contrast that with fusion detection, where, at least in my opinion, we haven't quite hit that point yet as a community in the bioinformatics world. This is an RNA based exploration, and I really want to use this one example

on the right just to highlight the challenges.

This was actually transcriptome sequencing from a castration-resistant prostate cancer tumor where we applied three different reasonably well validated and vetted fusion detection algorithms to this same upstream BAM and demonstrated a flurry of different results, 21 from this algorithm, 332 from this one, 131 from that one.

The nice thing is that the only overlap between all of them actually included the one fusion we really cared about, which was an MSH2. But I think this really emphasizes how hard this is analytically and how far we would have to go at this end of the spectrum.

I really used mutations and fusion detection at two different ends of the spectrum to point out that indels, copy number, some other types of approaches are somewhere in between. But to say that all of these methods are robustly validated and we can just call it a day may be a little bit farther than where we're at right now.

So figuring out what the variants are in our

next-gen sequencing panels, one challenge I think in some sense has been solved, but in other contexts it's still sort of a work in progress. Another big issue in the world of computational biology, next-gen sequencing, panel testing, and the clinic really revolves around tumor-only sequencing and how does one do that kind of analysis.

Tumor-only panel analysis is still I think -- there's a lot of conversation in the field about how we're going to do this and whether we're going to do this the right way because there's a lot of uncertainty about the false positive germline variant rate when one does a tumor-only sequencing because no matter what the strategy one uses, many germline polylmorphisms are rare and would not be filtered with any given database.

This, just to highlight this point, was actually work done out of Johns Hopkins from the Velculescu lab where they simulated this process with their filtering strategy by doing tumor-only calling. And depending on their approach, ended up generating between 30 and 60 percent germline false positives in

their tumor-only sequencing data.

So that not only creates uncertainty for false positive reporting, it creates uncertainty about when we actually find pathogenic germline variants in these cancer patients that we actually want to report, but we don't know whether they're somatic or germline. And I'm thinking about BRCA mutations.

Actually, Mike Berger, who was on the previous panel, was a member of a study that was reported at Memorial Sloan Kettering where they looked at this, the second citation on the bottom here, and really demonstrated that there's a lot of things that we can actually use the germline for that are relevant in the clinic. And even further, because so much of what this is based off of is actually germline databases and filtering from known germline SNPs, raises a question about whether there are disparities when we have germline databases that are predominantly populated by European ancestry people. What happens when we release this into the world where that's not actually the case?

To try to look into that, our lab is actually

investigating this in an in silico fashion. This is an example of looking at 157 clinical tumor-normal exomes where we achieved approximately 200X depths of sequencing in the tumors. And we tried to actually use this as a way to down-sample the exomes, model tumor-only calling in different conditions, and figure out whether we can come up with an optimal analytical pipeline.

That's actually what's sort of demonstrated on the bottom here, where we have sensitivity and positive predictive value curves for a representative 300-gene panel doing tumor-only sequencing. And on the X-axis here is a bunch of different approaches for filtering, using dbSNP, dbSNP plus XACT, dbSNP plus 1000 genomes, XACT plus Cosmic, all sorts of different flavors to try to find the right knobs to twist and turn to really kind of get us to that right sweet spot of minimizing the risks of reporting these false positives in tumor-only sequencing.

We came to the conclusion that, not surprisingly, using the largest germline database, so the XACT database that Daniel MacArthur collected as

part of a consortium effort, really sort of spanning worldwide data collection, along with recovery from SNPs and Cosmic and elsewhere and actually provided probably the most optimal approach, that still ended up yielding a 14 percent false positive rate, which I'll come back to in a moment.

The other aspect regarding disparities, which I think is another thing I would want to hit home and which is why any panel test that is doing this needs to document what they're doing, is that depending on the filtering strategy one uses, one is more or less likely to have false positives in non-Caucasian patients.

Just as an example, for this same 157 patients, we had patient self-reported ethnicities and ancestry, and in the non-Caucasian subset of these patients, we ended up having more false positives, germline false positives, when we just used dbSNP. And not surprisingly, when you go from dbSNP to a germline database that has over 60,000 germline exomes, that disparity is mitigated and in essence goes away. I think, again, this emphasizes things to think about analytically as we're releasing this technology into

the world.

I mentioned that we still had -- even when we try to optimize our strategy as best as possible, we still had a 14 percent germline false positive rate with this large 300-gene panel approach. But the other thing that we actually tested was what happens when we append a molecular pathologist at the end of that sequencing pipeline.

I am but a lonely medical oncologist, so I don't want to speak for molecular pathologists. But the cool thing was when we did that, the pathologists actually flagged a vast majority of these germline SNPs as likely germline. They get tiered in a bucket -- at least in this clinical lab setting -- that basically puts it at the bottom of the report in the unexpected section and, in essence, gets us down to really just one, two, or three different variants out of many hundreds of thousands that we actually have to worry about. And it's something that should be at least considered for these tumor-only panels.

If we started with all the different considerations regarding methods for different variant

detection, we looked at tumor-only sequencing -- the other thing I really wanted to touch on from a bioinformatics space because it also relates to what I care about in the research world, is inferring global genome properties from panels.

For that specifically, I'm talking about immuno-oncology because that's basically become the one and only conversation that we're having in the clinic for almost everything and every tumor type you can imagine. Much of that actually centers around inferring mutational load, neoantigen load, and all of those kinds of properties that one really can only infer -- in theory can only detect from genome-wide studies but actually try to use panel testing to infer that property because we think -- at least it's possible, in some patients who have mutational load regardless of tumor type and may be inclined to respond to these immuno-oncology, these new drugs, we want to be able to find these patients.

So for that same set of 157 patients where we did this down-sampling experiment, not surprisingly, when you have a reasonably large panel -- in this case

a 300-gene panel -- and you compare the true mutational load as inferred by the whole exome to this mutational load predicted by the 300-gene set, you end up with a pretty solid correlation, and that actually is useable.

In this case, it's referring to the tumor germline matched panel, but the same holds actually for tumor-only sequencing. But the ability of any given panel to infer global mutational load features decreases as your gene set gets smaller, which is not surprising. But I suspect because of the excitement in the clinic for these drugs, there's going to be a lot of interest in promoting one's panel for being able to predict these kinds of features. And whether or not they can actually do that is something we'd have to think about carefully when we're considering how we validate this.

I should point out -- I didn't put it in this deck, but not surprisingly, even though we can infer these genome-wide features, when it comes to actually identifying the actual immunogenic neoantigens, no panel is really sufficient because most of these neoantigens actually occur outside of any cancer gene

that we ever include in a panel test.

With that, I'll end. I think I sort have kept it to the 10 minutes. Thank you again for giving me this opportunity to present. And I think it's really exciting to engage with the FDA in this space because the field is happening -- is moving so fast, and the changes are happening before our eyes in the clinic. It's important that we all work together to get this going for our patients.

Panel 2 Discussion and Questions

DR. ROSCOE: Great. Thank you. I couldn't have asked for two better talks to launch this next panel. I would like to take a moment, though, to set the stage a little bit just to expand on this excellent platform, and that's that at the FDA, we have manufacturers coming to us wanting to market their oncopanels. They want to distribute kits. They want to sell them to labs.

So they say to us, "What do you need from us?" Well, we start with, "What's your intended use? What exactly are you intending to do with this test?" And then we go into looking at validation strategies

for our bread and butter sorts of studies. That's accuracy, precision, limit of detection, and specificity.

After that, there are some robustness studies, but those are really our core studies that we look at because we want to see what is the performance for these studies because that is what the manufacturer is going to go forward with, advertising their product. They're going to say this is the performance you can expect with our device.

So to that extent, we ask for a protocol. We ask for a protocol that begins from the specimen all the way through to the result because that is what helps the manufacturer get the best consistency, the best performance, in which they can go forward to the user and say this is the performance you can expect when you follow this protocol and run this kit this way.

However, a lab is free to do what they want. They are free to tweak. It is completely understandable that they need to modify the protocol, that they need to deal with each specimen and patient

situation uniquely. It's more about understanding that when you follow this protocol, I can expect this accuracy, precision, reproducibility, sensitivity, and specificity.

So with that, we'll go to the first set of questions because the first part that we struggle with is what exactly in this huge and complex field will enable us to have a very objective set of performance so that when customers such as labs and other -- this is certainly going to expand. People are going to be looking for more -- as this technology grows, people are going to be looking for more and more convenient and ready-made devices. We're not there obviously, but the technology is complex.

So it's absolutely critical for us to have an objective playing field for everybody to understand when I'm looking at these three devices, these three oncopanels, I'm comparing apples to apples, and I can make the best decision for my lab.

So starting with that, we always struggle. We know that we're going to have a representative variant approach. These intended uses are pretty

broad. We can detect SNVs, indels, small, large, copy number variations, translocations. So what do we need in order to begin -- what kind of specimens do we need and what are the generic, generally speaking, numbers that we need to be able to have an objective comparison of performance? What are the variants and considerations that should weigh in to determining what should be the representative variants for this performance evaluation?

So there are a number of considerations; clinically meaningful variants. Do we need to have samples that represent those specifically clinically meaningful variants? Do we have to have variants representative of rare diseases? Does it matter if the manufacturer's advertising 80 or 800? Should that impact the number?

Are there other -- we know about the challenging parameters. We know that we need to evaluate GC-rich content regions and all of these other global sequence contexts. What exactly should we be looking for in terms of beginning this objective assessment?

So here we have these experts. You all are experts in validating, and you know everything behind the curtain and in front of the curtain. So you bring your expertise in validating these assays. So I'll go ahead and invite anyone to step in and comment on what you believe should be a launching point for this type of validation.

DR. HEGDE: So I think this is probably one of the most critical and a sensitive question for the manufacturer of how much they need to go into in terms of selecting. Is it about the numbers or is it about the type? I think -- so it's my opinion that I don't think it's about the numbers. It's about if there is a variant that is included in the assay, it has been adequately tested for all the different parameters -- and by that I mean sensitivity, specificity, reproducibility, and robustness. If that is demonstrated adequately, I think that should be considered as good.

It's hard to hit a number. I mean, it's just simply not possible. The larger the number of variants in your assay, the higher the level of difficult of

getting all representative sample types and the mutation allele frequency that you're going to determine the sensitivity of the assay for. So it's a tricky question, but I think it's important that all the parameters that are usually assessed in a validation should be checked off in doing so.

DR. VAN ALLEN: Yes, I'd echo those points. I'd say it's hard to put a number of those, but I think you can imagine -- knowing that there are already very well validated clinically actionable alterations that fall into the buckets of cite point mutations, indels, copy number, fusions.

If your test is claiming you can detect all of those things, you have to at least start there, recognizing that the long tail of mutations in any given patient tracks to [indiscernible] zero doesn't quite get there. And I think you'll never be able to find all of them, but at least robustly being able to identify kinds of things so that you could always find the CDK for focal amplification or the EGFR indel I think is mission critical.

DR. DEIGNAN: I'm certainly not going to

comment on any numbers that I think we should use. But tying back to the previous panel, I think it's important to think back, when we were doing validations or when we are doing validations why do we have such high sample numbers? Is it because we want to try to generate all the different matrices? Is it because we don't have a representative number of variants in any one sample?

I think next-gen sequencing certainly changed all that. At least what I would like to propose is, in response to whether a variant needs to be clinically relevant for this representative variant approach, I would argue that clinical relevance is not that important in this context.

I would argue that when we're thinking about representative variants, we should use the power of the genome, the power of the exome, whatever you want to call it, that has a lot of intrinsic variation already: SNVs, indels, larger copy number changes, et cetera, because if we're going to claim that a particular sample has a change that we think we can detect, we're not necessarily saying that we can detect something to

the right of it or to the left of it, but I think that's sort of where we want to go.

One example that I would put forward are EGFR exon 19 deletions. If we think about validating that mutation, typically a lab would obtain, say, a single patient sample or a couple of patient samples, test that change, and then say, yes, we're able to pick up exon 19 deletions. But they may have only tested one version of that. There are obviously many different types of exon 19 deletions, so where do you draw the line as far as trying to figure out which samples you need to obtain in order to test that.

So like I said, I'm a proponent of you can now use fewer samples, fewer matrix variations, in order to get at good numbers to assess the analytic validity, especially the accuracy, for a next-gen sequencing somatic oncology test.

DR. EBERHARD: One of the things that I find interesting in the conversations are the background of what's the intended use of the assay. Is it specifically to detect particular variants in a targeted gene panel, for example, in the context of

selecting therapies and maybe look directly to those, or is it something much broader? Is it to find any alterations that might be there?

So I always have to keep in mind what it is that we're trying to accomplish with the assay, where do we want to go with that, and that sets some context. On one extreme, we may need to validate directly the ability to detect particulars or specific hot-spot mutations in well characterized oncogenes. Another may be the wide open approach, what can we detect as far as indels, SNVs, translocations, fusions.

Gene panels I find are often just that. They're called out by genes, and that's one way of bucketing variants, the genes of interest, and another is variant type, translocations, et cetera. So perhaps a middle ground thought might be that for a particular gene panel that we might at least go back to the huge data sets that are already in existence around what are the types of variations that have been detected and reported in that particular gene. One gene may have very different types of variations than another gene. So the particular variants to be addressed in a

validation might be gene specific, what are the things we would expect to see given the data that we have around genes.

Another thought that came from the previous discussion is the wonderful considerations of the absolute importance of tumor content and how we interpret particularly negative findings and how we need to validate those. Coming back to this arcane art of the pathologist estimation of tumor content and how do we do tumor enrichment, I was thinking about tumor types. For example, the Pan-Cancer discussion where immediately we think about, oh, different tumor types: lung cancer, brain cancer, breast cancer. But maybe another way of framing it is different architectural or morphological types.

Is it a cancer that tends to grow in a solid way well it's very easily sampled to have high tumor content or is it a cancer that tends to be diffusely infiltrative, and by it's nature will always have a low tumor content? Maybe that's one way of thinking about different types of cancers and what does that mean from our ability to gauge tumor content.

As a pathologist, what I would love to get away from is the subjective pathologist assessment of tumor content and to be able to imagine is there some type of an internal reference by which we can assess tumor content within a sample so we can know whether or not our negative calls are truly negative calls or if they're just below detection sensitivity limits.

One way of doing that can be by if I'm looking at KRAS mutation in lung cancer. If P53 variants are very high, then probably the negative KRAS result is a real KRAS negative, and it's not just due to low tumor content. One challenge I would put forth is are there ways that we could have internal types of controls by which we can judge the level of detection that we have in our samples.

DR. KLEES: So everyone made really good points before me, and I happen to agree with almost all of them, not to just be agreeing all the time. The types of samples that you're going to select initially is going to depend on the genes that are being targeted, and the hot-spot variants are going to be easier to obtain and easier to validate orthogonally.

So those are going to be included.

To the point that Dr. Deignan made, I think the variant types, especially with indels, they are going to behave differently, your medium size, your larger, your smaller ones. You need to have representation of all those included in the initial data set. And even with SNVs, an SNV in a repetitive region or a GC-rich region may not be as robust as in SNV in an easily targeted region. So I think you need to make sure that you include the good with the bad, initially.

DR. ROSCOE: Great. If I could just get a little more granular and ask about indels where we have different ranges and different sizes. So there are several different ways we could be looking at this. We could be asking that you demonstrate validity with specimens that have different -- in different bins, like zero to 10, 10 to 25, greater than 25 if the claim is, let's say, 80 or less, or we could say that you should have a indel, a single indel, in each of the targeted regions that you're interested.

Maybe we don't even need to be looking at

specimens. Maybe we could do some sort of walking the chromosome with an in silico analysis. What is the best way for us to get at indel performance?

DR. VAN ALLEN: I might can start the conversation, but I don't know if I can solve that for you. That's a tough question. What I would say is at least there are plenty --

DR. ROSCOE: Welcome to our world.

DR. VAN ALLEN: -- there are plenty of good examples of true positives, and I think at least starting there. I know, for instance, collaborating with folks at Brigham, I know that Brigham path folks have really emphasized making sure that you never miss the really key, well validated, clinically actionable variants. But I think the challenge then becomes what do you do when you step out of that space, and how do you know whether your assay is not missing increasingly important indels that may never have been described before.

I'm thinking mostly for my prostate cancer patients and somatic BRCA2 indels that we're starting to see more and more of in these advanced patients,

which has immediate therapeutic and diagnostic relevance. That's tougher, because I know from an analytical standpoint, like with the fusion detection algorithm point from earlier, different methods really yield widely different results. I think, at a minimum, being able to go back to the true positives that you know or you cannot miss -- the can't miss genes -- at least starting with there and demonstrating that is a good place to start.

DR. HEGDE: The way I look at indels is a little different from looking at single nucleotide changes. I think it goes back to the design of the assay itself and the methodology used for that, whether it's a hybridization based assay or amplification based assay, it's really the starting point for detection of an indel.

Also, especially in a hybridization based assay, you could probably design a synthetic probe for detection of a very large indel if you know the precise breakpoint versus a quantitative assessment of the reads to determine if there is a deletion of not.

So I think there are various ways to look at

it, but in terms of validating an assay like that, I think the important things to consider is how the assay is designed, are we doing a qualitative or a quantitative assessment to determine whether the indel is there or not, and then validating it with a set of true positive known samples to see if you can actually detect it.

I think in silico approaches are very, very useful here because you can come from the bioinformatic side and start validatiing from putting some in silico approach together and then go back to your biological assay, and then actually determine the sensitivity and the specificity of the assay itself.

DR. EBERHARD: One additional thought about the indels is I think our ability to call them well also depends on the local sequence context. So one obvious example would be a highly homopolymeric region would be much difficult to call than one that had more sequence complexity.

DR. DEIGNAN: And just to add another comment, yes, I completely agree that the local sequence context and the genomic region of a particular

variant is important. That's sort of what I was trying to get that, is because the normal genome has so much sequence variation across a variety of regions, a variety of chromosomes, a variety of sequence context, I think in this case, we can use that to our advantage.

Whether labs choose to then go after known hot spots I think is a second issue. But I think to focus on trying to get variants that are just indicative of those hot spots can be very limiting in terms of our ability to understand the performance of our NGS assay.

DR. KLEES: Other major issues that affect just indel detection is obviously sensitivity. Most NGSs I've seen, there's a significant higher sensitivity. So you need more reads present to be able to confidently detect the indels. Indel detection really seems to be highly affected by both the wet bench chemistry and even the bioinformatics process. So you have to make sure that you have the right tools at your disposal to be able to detect everything accurately.

DR. HEGDE: I just want to add one point

here, and I hope I say this correctly. I think there's a lot of focus on the depth of sequencing aspect for oncology assays and why it is so important when you're looking and comparing that to a germline detection assay.

The depth sequencing has to be looked at in what has been just discussed from two aspects. The qualitative aspect is a neighboring sequence complexity irrespective of whether you are detecting a single nucleotide variation or an indel. That's one way to look at it, and then the depth of sequencing of minimum cutoff that has to be put in place. So you're actually going to detect what you're looking for. So I think both aspects have to be addressed when you are putting something together and actually validating it.

DR. ROSCOE: All right. Great. Just to tie this up, I think I can probably anticipate the answer to the next question and the last question in this section, which is for limit of detection. In terms of reporting out a limit of detection performance, that's all over the place with different variants. But obviously, manufacturers like to go and commercialize

limit of detection claims because that's a very critical feature for these devices.

So where do you all stand when it comes to actually making a broad performance statement about a panel when it comes to the limit of detection having been done with a few variants in each variant class? It sounds like you would probably not be in favor of that but maybe reporting by the variants.

Do I glean that from your discussion point so far? Go ahead.

DR. EBERHARD: One thought would be that the LoD has to start with a specific variant to be evaluated. And if a claim is made more broadly, then at least initial data has to be done prospectively showing that in fact your claim can be prospectively generalized to new variants. And if that is done a sufficient number of times, however many that is and it can be demonstrated to be reliably generalizable, then we could accept that.

DR. KLEES: I agree that the sensitivity, you just can't be one and done. Different genes behave differently, even just different areas within different

genes. One exon may be harder to sequence than the other. So it's really better to have a representation if you can, just keep diluting your samples down or even confirming something that you have that are near your LoD just to prove that they are in fact true positives and that your sensitivity is strong for that particular variant type.

DR. ROSCOE: Okay. Great. So let's move on to the next section, set of questions here. Let's just say we've got our variant panel. We're ready to move forward with the studies. One of the topics that we struggle with is what is truth. This actually ties back to our troubles with determining what is an adequate representative set because it's very challenging then to come back with the truth on that set.

In your experience, what are acceptable orthogonal methods for accuracy, and what consideration should go into confirming results so as to avoid bias? In the December workshop, there was a lot of discussion about systemic bias incorporated in due to just looking at the positives that you find. In our world, we would

probably ask for some confirmation of randomly selected negatives as well.

If you could just comment on what you believe this to be, what you believe would be the most efficient and effective way to look at accuracy by a comparator?

DR. DEIGNAN: Well, I would certainly say that it's not Sanger sequencing. You could make an argument that it should be an alternate next-gen sequencing approach, but I think a lot of people would disagree with that. My personal opinion is that the only thing that you can compare your new data set to, to be certain of its gold standard validity, is some sample with truth variant calls. And how that sample arrives at truth variant calls I think is something that a lot of groups are working on right now. But as far as a comparator method, I haven't seen a good one.

DR. HEGDE: I would agree with Josh on that, definitely not Sanger sequencing. The claim of detecting a very low allele frequency has to be taken into the context of what the sample type is, what you have actually analyzed, and what controls you run in

your assay itself.

Probably a cross-comparison across the different runs that have been done, and if you're able to detect the same allele frequency again and again in the same sample type, it probably has more value than trying to go outside it and trying to find a method which will give you -- which can be consistently used to confirm the accuracy of the detection of the variant itself. This will get even more complicated as we start looking for indels because there's just no method that will give you a definite answer.

DR. KLEES: So I guess I'm not as against Sanger as everyone else. I do believe Sanger has a place being the gold standard and all. And especially if a lab that's performing this has it established and set up, it's very easy for them to confirm the accuracy that way. But I think you can confirm the accuracy by comparing it to another lab performing NGS.

If you do panel versus panel, you may only have three genes that are positive but the rest are still going to confirm negative. Also, if you sequence well characterized cell lines, you're essentially

getting all the negative regions that everyone else got. So I think you can confirm negatives that way, too.

DR. EBERHARD: So in considering this, I was thinking of accuracy in what. And there's accuracy in mutation detection and there's accuracy in mutation calling as Dr. Pfeifer talked about previously, and that those can be separated perhaps into two different exercises as well as concepts and different approaches in how we might determine those.

For example, Dr. Pfeifer talked about the use of engineered plasmid constructs in mutation calling, so that type of a construct might be appropriate for mutation calling, or even in silico might be appropriate for that. If it's accuracy in detecting mutations from particular samples of interest, then we might need to use a different approach.

So I think that we want to be clear when we're discussing exactly what it is that we're discussing so that we can better get our head around the problem.

DR. ROSCOE: All right. In terms of the

actual performance reporting, in terms of sensitivity and specificity -- in the last panel, it was interesting because we heard a lot of perspectives on how that reporting is done. In your experience, how do you think sensitivity and specificity should be calculated and reported?

Is it enough to be overall or should it be on a more granular basis, by targeted region, by variant type, and should performance metrics be reported in conjunction with this type of performance in terms of did you meet some quality metric, which regions met this quality metric? These are the limitations. These are the error rates.

Basically, speak a little bit for us about how performance should be reported in terms of the assay.

DR. VAN ALLEN: Well, I guess I'd just start it off by stating that all I can speak to is what we're doing locally, but the first part of our report really is sort of a global quality assessment, global coverage, purity estimates by both pathology and by in silico approaches, and I think what we're trying to

now do, recognizing that the end consumer doesn't oftentimes understand a lot of these technical features. I'm speaking about my colleagues, so I apologize in advance for that. And the clinics. Sorry. That was a joke.

(Laughter.)

DR. VAN ALLEN: We try to actually make it as straightforward as possible to recognize when, whatever your sensitivity/specificity is for any given variant caller you have, there are certain global properties that are going to limit or restrict your ability to make any claims on that for any given sample.

At least, if you start out by just reporting that, which isn't always the case, and then report it in a way that the consumer can understand what that actually means and how that impacts the rest of the downstream analysis, that goes a long way towards making the test useful.

DR. EBERHARD: Part of the conversation in the previous panel was related to the practice of medicine and how that plays into certainly the pre-analytical part of our test. Actually, it extends all

the way through our next-gen testing, first with pre-analytics, but then after the data comes out -- for example, we have a molecular tumor board that sits and looks at all the data in a pretty granular fashion, including quality metrics, et cetera. And we make an objective judgment from those metrics as to what we decide to go home with.

I thought that Dara, for example, in the last session gave a great example of a sample that had a whole bunch of -- you know, 3, 4 percent VAF changes that were artifact, and then there was a bona fide mutation that was 13 percent. So we could eyeball those as a molecular tumor board, all of us sitting there, and say, oh, okay, we know what to believe and what not to believe.

So our question is, is what parts of the assay performance and interpretation and utilization should be left to the art and practice of medicine versus which can be controlled and well defined and included as part of the assay itself so that myself as a pathologist, I don't have to worry about what's going on inside the confines of that assay because I'm

worried about things that are going on outside of it.

From that standpoint, are there ways that we can look at what molecular tumor boards do and incorporate that if there's some type of a systematic assessment, incorporate that into part of the package and quality metrics that are included in an assay.

DR. VAN ALLEN: We actually do something similar with a molecular tumor board and I think even emphasizing the idea of making sure that a molecular pathologist is at the end of our clinical assay. But ours is a one -- it's not a scalable solution. I think what we're struggling with is -- so I agree with you, and I think it's mission critical for actually doing the interpretation and doing it correctly for now, but we're struggling with how do we scale that when recognizing that we can't -- there are places we can make these molecular tumor boards work, and then what do you do for the larger community.

I'd be curious to get your thoughts and everyone else's because I think it's really an important part of this.

DR. EBERHARD: I agree it's not scalable for

a number of reasons. One is the number of cases it might take, the type of people in personnel that it takes to do the analysis. And another is that the evolution and accumulation of new data happens so rapidly that as human beings, it's extremely difficult to keep on top of the literature. So this is where artificial intelligence or some type of bioinformatics solution for encompassing that and keeping it up to date is incredibly important.

DR. ROSCOE: All right. Thank you. So moving on, we had a lot in the next talks as well about somatic versus germline variants. So talk a little bit about that because these panels do detect both in terms of whether or not the performance should be reported out separately and if there are different strategies for validating them, and any drawbacks that might be encountered.

DR. VAN ALLEN: I covered some part of it in my presentation earlier, but I think given the challenges of doing somatic-only testing, I think actually it's critical to make sure that if that's the test one is doing, that there is a strategy to at least

be able to report or figure out what your germline variant calling is within your given pipeline however you do it, and that that false positive rate, or whatever you want to call it, is part of what you're claiming and how you're doing your process.

My concerns are end users, where oftentimes there's not clear distinction or a clear understanding by the clinicians about what that even means. I'm not clear whose responsibility it is to really make sure that that's understood, but we can at least -- there are many ways to make sure. You can at least know how your test performs. You just run a tumor-only -- you start from the beginning, do tumor only, and then tumor matched, and do it in silico and everything in between. There are lots of ways to figure out how your assay performs in that space.

DR. HEGDE: I think it goes back to the question of how the validation is being done by the manufacturer. Are you using tumor-only specimens for the validation? I would say that that should not be the only type of specimen that is used for validation. There should be normal tumor pairs used for validation

so it can actually show what allele frequency you are differentiating between a somatic versus a germline variant.

Also, that goes on to the bioinformatic pipeline of at what level you are going to put your filters in because you could easily filter out something which is important as you lock down this assay to say that you're going to detect only germline or somatic.

So it's a tricky question, but I think in the validation process of showing the performance of the assay, this could be addressed at that level to include what type of samples to do the validation itself.

DR. DEIGNAN: So I'll say that our somatic experience so far has just been with tumor-only sequencing, which I know is probably what a lot of other people do, too. So we've certainly been thinking about this. I think one of the challenges I certainly have is when you try to do tumor normal comparisons is deciding what is normal because, obviously, you could have those same somatic changes in what you classify or somebody classifies in your group as normal tissue.

Even if you use a blood comparator, nothing says that a particular somatic variant isn't floating around in the blood somewhere, which might then result in you ruling it out and potentially not giving a potential treatment to a patient, for example.

So I think it's just important to keep this in mind. One thing that I was thinking about with the somatic normal comparison, too, is that we've evolved and become better with the pre-analytical steps, and there was a lot of discussion earlier about pathologists review of samples and macro- versus microdissection. If you have a particular sample and you say don't macro- or don't microdissect, and you get a 10 percent allele frequency, you can pretty much say that it's a somatic change; whereas if you do macrodissect, and now all of a sudden that allele frequency becomes around 50 percent, it obviously makes your life more challenging.

So I think if we're going to try to minimize the number of germline variants, some form of germline comparison I think is going to be necessary. Obviously, the second talk talked a lot about the

factors which might complicate that. But there is an inherent challenge with doing somatic-only sequencing. And especially I think if you go beyond these known mutation hot spots, which is what people are now going into, you end up with all these germline challenges, which the germline community is already trying to understand and address.

DR. HEGDE: I just want to add a little bit to that. I think when you're doing tumor-only validation, the problem is that if you've not done some sort of a normal comparison -- and again, I say that the normal tissue is something which is absolutely done -- the pathologist has already validated it to be a normal assay. You still have the risk.

But the problem here is that I have seen this in my own lab, that if a genetic counselor is involved and a report has come back saying that this variant is present at a 50/50 allele frequency, the genetic counselor will immediately pick up on that and request a confirmation for germline.

I think the manufacturers -- and to do that validation up front and show what the assay can do and

cannot do is extremely important because that translates over to the bioinformatic pipeline and the report generation from the labs. So they need to know what they can see on the report or cannot see on the report.

DR. ROSCOE: Thanks. I've run out of time on this session, but I do want to devote just a few more minutes to the last two questions because they're so critical. And that is, how reasonable is it to allow a lot of this validation to occur with actually BAM files, and should process controls be integrated into this validation?

Normally, we actually don't require genetic tests to come in with controls. This sort of situation might be a laboratory function in terms of the replicate testing, the dual strand synthesis, or normal matched blood. But in some ways, it can also be used to support the validation in terms of truth for the obtained data.

So can you speak about incorporating BAM files or FASTQ files from that starting point in the analytical validation and whether or not we should

implement process controls, different types of controls in the validation strategy? I stumped everybody.

(Laughter.)

DR. HEGDE: So you're talking dry bench validation? Is that what you're saying, that there is no wet bench validation done in a manufacturer's statement, only the BAM file with different versions or different allele frequencies of a particular mutation to go through the validation process? Is that what we're talking about?

(Dr. Roscoe nods in the affirmative.)

DR. VAN ALLEN: I think there are a lot of fun ways to make like Franken-BAMS [ph] and do all sorts of fun things on the computational side. And I spend a lot of my nights and weekends doing that for fun. But based off of the first panel and what this panel has discussed, I think there's so many factors that play for any given assay that are before that. Just having that, if that's where we're going, would make me a little bit anxious.

DR. KLEES: I mean, I like the idea of the Franken-BAM file, but I think it has to complement the

validation of the entire assay. And the only that I can think of only using a BAM file of FASTQ file was if there was some minor modification to the variant callers that they were using such that none of the wet bench chemistry was affected. Then, yes, you could potentially just use previous data or Franken-BAM files to see if it still calls it the way it should.

DR. ROSCOE: All right. Great. Let's move to the final question, validating modifications to the panel. So now we've successfully validated the device. We have our performance. What is needed to actually come in with changes to the panel? Is it reasonable to have a change control process in place?

Should we actually look at more -- not necessarily request different specimen validation but look at granular characteristics such as all of the QC characteristics that after each step, the post-alignment, the mapping, these sort of variables, quality scores, overall, needing some sort of threshold in that sense, or do we need to actually request additional validation?

What are the minor panel modifications that

you could say do not diminish assay performance? So where is there wiggle room here in terms of when a sponsor modifies a panel and also in terms of actually wanting to sell a similar panel? So let's say they validate an 80 variant panel and they want to sell a smaller panel.

DR. KLEES: I'm not sure you can determine diminished performance until you actually run it. Any wet bench modification I think would have to be analyzed, not to the extent of the original validation, but you would have to take samples, preferably ones that you did in the original validation and make sure that there's not a significant change in coverage and sensitivity, reproducibility, things like that, so that nothing was negatively affected.

DR. EBERHARD: I find that a phrase like what types of changes are not expected to change performance characteristics is kind of scary because I've known some pretty optimistic people in my time, and sometimes we get surprised and things happen that we don't expect in the laboratory, and that's the exact reason why we do validation.

In considering the questions that we're discussing today, one of the overarching things that I've felt is should we or should we not basically change our approach to validation because we have next-gen sequencing to deal with now. My initial reaction is that our reasons and philosophies and approaches for doing validations aren't changed by the technology. They're technology independent.

So we shouldn't give up good, hard, critical, solid science just because we've got a lot of data to deal with. What we need to do is to understand how to draw the lines around and define what it is exactly that we're doing in order to develop appropriate validation approaches. So that was just kind of a high level thought.

DR. HEGDE: So I'm not very good with these type of questions where what type of changes are not expected to change performance characteristics. But I think one way to think of this is that there are exceptions that happen all the time in the lab. And none of us can say that this will not affect the overall performance of the assay.

But in general, looking at the different types of variants targets or even just a change in assay reagents, I think once an assay is locked down, you would expect it to perform the same way every single time. And if you have controls included in the assay itself, if there is a variation in the performance of the control that has been included in the assay, that could indicate something that has changed. But again, there are changes in rare exceptions that happen all the time in the lab.

DR. DEIGNAN: Just to speak to the topic of controls, again, I'm a proponent of some of the stuff that was covered in the previous session about using quality control metrics from the sequencing step as really our new positive control and not running any sort of separate control.

We've been doing this for a number of years already with Sanger sequencing where we haven't necessarily run a positive mutation or variant control with a regional or even necessarily full-gene Sanger sequencing assay. We've just used the fluorescence intensity and the fact that our sequence matches what

is "our normal germline" as evidence that we're faithfully sequencing the entirety of that region.

So I think it's a logical extension to translate that to a next-gen sequencing approach. So using whatever we define as a particular set of quality control metrics, then any minor deviations that could occur from day-to-day, sample-to-sample or things which are outside of our control, we would be able to see whether they would have an impact on downstream assay performance as long as those quality metrics were set up appropriately.

DR. HEGDE: I think this also brings up one -- I'm just trying to think. The allele panel -- and Dara brought up a very good point, that if the sample quality is not optimal, or if there is some sort of a compromised starting point where you don't have -- it depends on the cellularity or the tumor content, do you not do the assay because it has been validated in a certain setting?

I think that's where we're starting to get into what the physician should do for that patient because probably doing that assay might help that

patient or may not help. It's kind of a hard thing to get to.

DR. EBERHARD: Along that line, that can go along with how we use our quality metrics. And one use would be to say, no, we're not going to take that sample because it's not high enough quality, or we'll say, okay, we'll run that sample; we'll see what we get. But at least we do understand the quality limitation so that when we get the data interpreted, then we can do it in an appropriately cautious context.

DR. ROSCOE: Would anyone like to comment on anything we missed? We had a huge wish list when we were devising these questions. And then we thought, well, let's just narrow this down for the foundation types of questions that we have right now. We thank you so much for sharing your expertise with us. You've certainly given us that foundation today.

Is there anything you would like to comment on that we haven't touched -- anything you would like to add that we may have missed?

(No response.)

DR. ROSCOE: All right, then. At this time,

I'll invite members of the audience to come ask the questions. Any questions?

DR. AISNER: Sorry if I look familiar. I have reservations about this concept of using variant allele frequency to make implicit decisions about germline status. I've seen certainly examples of polymorphisms that exist far below 50 percent because of aneupolidy in the tumor cells, which have diluted the SNP down.

I think just because something exists at 50 percent doesn't make it a SNP, and just because something doesn't exist at 50 percent doesn't make it a SNP. And I think we have to be very, very cautious about using variant allele frequency to make those assumptions or draw those conclusions.

I think that the presentation that Eli presented showing that once you take a set of unknowns and present it to people who have the right databases, the right tool sets in front of them to make the informed curation decision, that's really where we need to be thinking about for looking at how do we decided what's a germline SNP, what's possibly a germline SNP.

The next question really is a reporting question. You have something that you go, okay, this is definitely a SNP. We're not going to report that. But then as you refer to in your practice, you've got tier 1 and tier 2. You do have some of the things. You go, well, gee, an XACT has a general population frequency of 0.002 percent, so it could be a private or quasi-private SNP. Do I report it or do I not report it?

There are also opportunities to use language in your report to indicate the possibility that something is a SNP, and I think that we're losing some of those shades of gray when we talk about this in absolutism of this is a SNP, this isn't a SNP. We need to use germline tissue or non-tumor tissue. I think there are shades of gray in there that allow for a professional interpretation to make those informed decisions.

DR. VAN ALLEN: So I'd echo those points. I agree, and I think we've been burned plenty of times with examples like the ones you're describing where you actually will have -- it's pathogenic alteration in

let's say BRCA2 or P53 that could actually either be germline or somatic and has the same function and downstream effect, but you can't actually tell the difference based off of the sequencing unless you have the matched germline sample.

Like you also stated, this actually creates a good opportunity for innovative ways to actually report this information recognizing that there are shades of gray, that there are no absolutes, but how do you actually convey this complex knowledge to the folks who are trying to make clinical decisions from this. I think that's a problem maybe sort of downstream of this panel's considerations, but it's almost more like clinical informatics user interface kind of representation problem.

But I think it's a really important one because our experience has been the clinical oncologists who are making the decisions off of these tests need help. And I think especially for -- we're trying to distinguish somatic and germline as sort of a key principle that we recognize as being really important, but they may not understand what those words

mean. We have a long way to go.

DR. HEGDE: I just want to add one thing. I think the use of XACT -- we've been talking about how useful XACT has been in the clinical setting. But one thing to remember is that when you look at a particular variant in XACT, it's really important to look at how many alleles have been reported because you can very easily make an error in determining whether this is a false positive or false negative report.

In XACT, also there are issues there that you have -- in my own lab, we use a minimum of 10 alleles or more as a limit. You cannot just assume that XACT has the right allele frequency or the right number of alleles in the homozygous and heterozygous state to make clinical decisions.

Do you agree with that?

DR. VAN ALLEN: Yes, I wholeheartedly agree. I think you've also seen the other situations where you have, again, pathogenic germline alterations that are in places like XACT or dbSNP, and then you filter out those things, and what's in there? Things like KRAS, G12V, and other things that are actually obviously very

important somatically. And if you don't very carefully consider how you construct your strategy for dealing with this, like you state, you may end up in trouble one way or the other.

DR. HEGDE: I think that's why the concept of variant allele frequency is very important in relation to the clinical interpretation of the assay itself, and then comparing it to that frequency in the different databases.

DR. ROSCOE: The man in the back?

MALE AUDIENCE MEMBER: Yes. I'd like to come back to this slide that's up now and get the panel to think about a specific example. So say a vendor has a focused panel with actionable genes where there are therapies targeting variants within the genes. And then a study gets published, and there's a new gene they want to add.

I think that's what gets to this, is if you already have your panel and say you're doing a hybridization capture approach and you want to add a new gene to that panel, I think the FDA's looking for guidance, and I'd like to hear the panel's impression.

Specifically, I think Robert and I know, Dave, you've been running a panel. How do you go about assessing the performance of the panel with the addition of new genes, and what do you think should be the standards that labs are held to for that process?

DR. KLEES: So how we typically look at it with New York, if someone wants to add a new gene to an already established or approved panel, kind of the minimum things that we expect to see is the accuracy specificity of the new targeted region, how well do you capture it, how well do you sequence it, and also showing that the performance characteristics of all the other areas that you're targeting aren't negatively affected. Those are at least the basic things that we look for to start off, and then you could go on to validate that you can detect whether if you're looking for indels, SNVs, things like that to go for the accuracy of the variant detection.

DR. EBERHARD: I think that it also depends on what exactly adding more genes means technically, whether that means a redesign. So for example, in a bait/capture approach, which we use in our laboratory,

does that mean a redesign of all of the baits or does that mean dumping a few new baits in with the old soup and adding it on. Even in the latter case where it would seem as though the formally used baits should not be affected, we still don't know that for sure, so we want to run some type of a confirmation to ensure that.

I personally don't feel that probably we would need to do a full demonstration of the ability to detect various specific mutations across the entire gene panel, but at the level of on-target coverage, the number of reads we're getting for each of our targeted sequences, et cetera, et cetera, those types of basic quality metrics maybe even to the point on Illumina platform, I'm looking at clustering to ensure that clustering hasn't been affected. So those types of QC metrics certainly should be examined across the entire panel to make sure that we're maintaining the basic QC performance that we're expecting from the panel.

DR. DEIGNAN: And I guess I also wanted to ask, when I was initially reading this question, I was also thinking about the example where the panel did capture or did sequence a particular gene. And then it

wasn't previously analyzed or reported, and now the vendor wants to make it analyzable and reportable.

Was that also a consideration with this question? Because there I think you're talking about something different. I think if, yes, you are adding targets or adding something to the wet lab part of it, then I agree you do need to do some form of validation because what you don't know you don't know, and it's good to know that. But I think if you're just doing sort of an informatic expansion of the windows so to speak and you're not actually changing any wet lab bench work, then I don't think you need to do any additional samples per se, or even necessarily -- you would probably just do a bioinformatic validation at that time.

DR. HEGDE: I would just add one more thing. Another different way, a different strategy, is to look at this. It depends on the -- if there is a significant publication which says that this could really affect the treatment of a particular type of cancer, then we might have to consider some deviation strategy so that the manufacturer can do some panel

studies and immediately get that new target. Now, it depends on the number of targets, too.

The other way to look at it, and we have done this in my lab, is to allow some spiking strategies where you can spike in a separate -- the target, the reagent, the bait into your original validated reagent so that you can continue offering the assay while there are some approvals that have been put in place.

DR. ROSCOE: Dr. Klein?

DR. KLEIN: Roger Klein, Cleveland Clinic. My question is for Dr. Van Allen. So you presented a slide showing mutational load in the end in an effort to develop an assay that would be predictive for responsiveness to, for example, PD-L1 inhibitors. And you suggested that commonly mutated genes, which would contain mostly driver mutations wouldn't really be the likely source of mutations that would predict antigenicity.

What I'm wondering is, in your research, have you been able to find regions or specific genes or areas basically which you could hone down on and eliminate the noise so that you could obtain greater

predictability?

DR. VAN ALLEN: The short answer to your question is no. It's actually a really interesting open question in the field. And just to take a step back for the group who may not be following this area of clinical oncology as closely, the ability to stratify patients based off of a mutational load for CHK1 inhibitor responses is sort of a provocative idea. It hasn't been prospectively proven, but it's an interesting concept.

The idea here is that the more mutations you have, whether they're driver mutations or just passenger mutations in the tumor, the more likely you are to have your tumor accidentally make mutagenic neoantigens or peptides from those mutations. And then the more of those you have, when you give the CHK1, they'll respond.

The challenge is that when you're testing let's say 300 genes, but these neoantigens can occur in any of 19,700 genes, we just don't know, you're going to be missing -- if you actually want to find those specific exact neoantigens because let's say you want

to make a vaccine against them, which there are trials happening to do just that, that are ongoing now, you're not going to be able to find them with that strategy.

The biggest limitation of doing what you're proposing is that most of the neoantigens we're predicting are based off of in silico models, things like MHC, because one has not been able to do that ultimate experiment of making all theoretical let's say 9 and 10 amino acid peptides against all possible HLA types. That would cost a lot of money to say the least.

So I think until that happens or until we have more knowledge about which are the peptides that experimentally actually provoke response or that are seen in the patients who have responses, or when one makes the vaccine that leads to the response, it's going to be hard to be able to say, you know -- most of these neoantigens are occurring in let's say not these 300 genes but maybe another 400 genes or 500 genes and go that way, but I think we're not quite there yet. But that would be a fantastic way -- if we can solve that, that would actually be a pretty big step forward.

DR. KLEIN: Thanks.

DR. ROSCOE: All right. Well, thank you so much. Let's give our panel a hand.

(Applause.)

DR. ROSCOE: We appreciate this. You've definitely given us a lot of useful information.

Now, we will break for lunch. We will reconvene at 1:30. You can get lunch at that kiosk out there.

(Whereupon, at 12:31 p.m., a lunch recess was taken.)

A F T E R N O O N S E S S I O N

(1:30 p.m.)

Panel 3 - Abraham Tzou

DR. TZOU: So hopefully everyone enjoyed their lunch break and the break for whoever is watching on the webcast. My name is Abraham Tzou. I'm a medical officer in the Division of Molecular Genetics and Pathology, and I'll be moderating this third session on clinical and follow-on companion diagnostic claims.

The format, we'll start off with a couple of presentations. I'll have the presenters briefly introduce themselves so they can describe their perspective and background in the area, and then we'll continue with the other panelists introducing themselves and go through some of the questions and discussion topics.

Our first presenter is Dr. Shashi Kulkarni from Wash U, and I'll have him come up.

Presentation - Shashi Kulkarni

DR. KULKARNI: Good afternoon. It's a pleasure to be here. As a way of background, I'm a professor at Washington University School of Medicine in St. Louis, and I also run the clinical lab, which does all sorts of clinical genomic testing:

cytogenetics, FISH, microarrays, and NGS. And you've heard this morning from Dr. John Pfeifer in the first panel about the different approaches in the pre-analytical phase, so I'm not going to go there. And we have had a wonderful introduction to the other aspects of the analytical phase.

So what I've been asked to do today is to walk you through the process we go in a clinical lab for determining the pathogenicity or clinical significance of the variants we see. What I'm going to do is share with you the process we go through, then share what are the challenges, which are still remaining here in clinical utility and determination. And I'll end up with a positive note and share with you a lot of new efforts, ongoing efforts, because this is a work in progress, and it takes a village to come up with enough evidence to determine the clinical actionability and clinical utility.

Here are the list of my disclosures. On the top, as John Pfeifer showed this morning, I'm a professor, and I'm the director of the lab, which generates revenue. As I said, determination of

clinical significance of variants is the task I was handed over.

I want to start with going back to basics. Variant issues are not new. I'm trained as -- I'm a clinical cytogeneticist and a clinical molecular geneticist. Since the last 50 years -- and I have to show a karyotype because I'm a cytogeneticist. We have been faced with variants -- shown here is chromosome 9 variant, which is benign.

We see this all the time, not only in germline situations, constitutional analysis, but also in cancer where we see this. And now we know that these are benign variants; whereas we see additional marker chromosomes, and we go through a tremendous process, which includes looking at literature and doing parenteral studies and constitutional and cancer association studies. So we have enough information and experience on the determination of the pathogenicity of these variants.

I also wanted to share with you a variant rating system, and there are many such systems. This is one of the variant systems, which was published, and

I'm involved in this. I just wanted to share this figure from that paper where it nicely shows the thought process we go, whether it's a copy number variant or it's a single nucleotide variant. Essentially, you go through this process for all different classes of variants.

This is the current thought process we go through for NGS since this an oncology focused meeting. This is the current thought process we go through, whether this variant leads to change in clinical management of the patients. And we have literature for these tier 1 variants; whether the variant predicts survival or other clinical endpoints independent of any specific treatment, whether it's a prognostic variant; then can we inform therapy or guide therapy based on a particular variant. There are several examples out there.

Finally, looking at these variants, assessment of these variants to monitor the therapy. I would not talk a lot about the last one, which is the pharmacogenomic because it's still not yet part of our clinical practice.

We have lots of external databases which we use and tools to understand the database and to look at the clinical significance of these variants. I just listed the different databases we use everyday to interpret the sequence variants. The list goes on. I have all these URLs. If anybody's interested, I'm happy to share.

Then there are algorithms which we use to predict the functional impact of sequence variants. There are lots of caveats about using these. I'm not going to go in technical details. Maybe we could discuss this while we are having our discussions. But there are many of these, and one has to be very careful on the significance of using one versus the other. But we have enough information now and some nice tools, which also have statistical calculations to help us out.

I've put this slide again here just to belabor the point that this is Wash U's somatic variant classification scheme. And I also want to mention here, unlike the constitutional or inherited NGS testing, there is no guideline as of yet for this

variant classification. However, the Association of Molecular Pathology, American College of Medical Genetics, and ASCO have a joint committee. I'm also part of that which is now in the final stages of coming up with these guidelines, similar to the constitutional guidelines, which were published last year. These guidelines should be out in a few months.

So this is how we report our variants. Level 1 is predictive or diagnostic like I showed before. For example is BRAF V600E. There's no discussion. We know there's enough information. Level 2 is prognostic or predictive in other tumors. And this is an important advantage of next-generation sequencing. Before this technology, we were not looking for the variants which were not known to be in a cancer or for different origin.

Here is an example of the IDH1 gene. R132 variant is found in acute myeloid leukemia, and there are IDH1 inhibitors. Well, what do you do if you find this variant in colon cancer? So we still want to classify this variant as an important predictor of prognostics or even with therapeutic implications. But

because it's not very clear yet, we would put that as a level 2.

Level 3 is pretty broad where we see this in Cosmic data or TCGA data, where we don't have a lot of clinical outcomes associated with it, and it's not very clear as to how this would be clinically significant. But we do want to mention these variants in the report because this classification is a dynamic process. This is not a static process. With two, three months down the line, there might be a big study from cooperative groups, which might change that classification. And the level 4 and level 5 are known, variants of a known significance or polymorphisms.

So we are at a stage in cancer where we are witnessing a paradigm shift. I don't need to preach to the choir. But we have challenges. We have huge challenges. These databases are not accurate all the time. There are cell line data, and we all know that cell lines have inherent issues where just because of the growth advantage in the culture, they gain antibodies or they have mutations, which are more like cultural artifacts.

So we need clinical grade variant, knowledge based, and it's very important. In constitutional inherited genetics, NIH has been doing a lot of work with this ClinGen, which is multi-institutional, a multi-NIH funded grant. So we wanted to see how we can use this ClinGen approach, which is shown very nicely in this figure, which is published in the New England Journal of Medicine, where it talks about clinical validity, pathogenicity, and utility.

The center of all of this sits the patient. I mean, how can we improve patient care through these amazing tools we have? We wanted to start thinking about wearing the success we have had in the germline, so we established a somatic work group. And this is the vision statement of that work group, which the main goal is to develop a process that supports clinical grade, a determination of clinical relevance to be used by physicians, labs, researchers, and guideline developing groups.

I've listed the mission statement, which is that we want to develop standards of these classifications and create this database, which can be

used by various other groups to enable precision medicine.

I'm not going to show the details here, but we have an amazing group of people who come from bioinformatics and a clinical lab like me. We have oncologists in the group, and we have over 50 people involved on a day-to-day basis. And some of them are in this room right now. So it's been going on very well. We have been working -- so this is another major issue we have. There are several other groups working on this. We don't have a common platform where we can work all together. So we have been able to successfully form collaborations and pool our resources so that we can work in unison.

One of the major achievements we had -- because most of these variant databases don't even have the same syntax, and how we gather the data is very, very important. So we created these minimum variant level data requirements and case level data requirements, which are listed here. I don't expect you to see this. But I was amazed that nobody has actually gone and standardized this data collection

approach.

So we're working with NCBI ClinVar project to make sure that whenever we start gathering data from all these clinical labs -- there's already huge data sets available in clinical labs. So we want to harness those, but we want to harness them in a very defined, structured way so that we can understand this much better.

So we're working with several collaborations. As I said, AMP guidelines for variant classification should be coming out anytime. We have lots of future activities coming on line. We've had several meetings face to face, and we're creating more approaches to do this assessment better.

So in the end, it takes a village, as you can see here. And this is an old slide, and I would say I will need three more slides to fit in all the people. So we've been very fortunate. I wanted to share this with you so that we all understand that there are efforts which are ongoing and it's not all a lost cause. Thank you.

(Applause.)

DR. TZOU: Thanks, Shashi.

Our second presenter is Dr. Dane Dickson, and I'll ask him to come up.

Presentation - Dane Dickson

DR. DICKSON: My name is Dane Dickson. I am a medical oncologist. I am not a molecular pathologist, nor am I a molecular oncologist. I'm the token user of these tests that sees patients. I'm also the CEO of a non-profit volunteer organization known as the Molecular Evidence Development Consortium that's trying to truly work together to try to unlock many of these things.

I'm going to simplify this a little bit. I like to think in simple language, and I think it's important to get back to basics. Some of the essential tenets that we all understand from when we were in junior high science is the scientific theory. You come up with an observation.

We think we have something that really can make a difference in the treatment of patients. We've got new technology, whatever that observation is. We ask questions, can this help someone? And then we form

a hypothesis, yes, it can. And then we do tests to either show it can or it cannot.

If you start to really take the hood off of where we sit with this brand new technology of next-generation sequencing, we really have jumped the gun a little bit. We've made the observations. We've asked the questions. You'll notice I've said NGS is the Swiss army knife of testing, and it's either a question mark, a period, or an exclamation point depending on what you're looking at.

It may possibly do the work of Sanger, IHC, FISH, PCR, and I'm sure it can do your laundry in certain applications. It probably needs less tissue, and it may be at a lower overall cost than doing sequential biopsies on patients. The hypothesis, NGS can replace and it may be better than other testing. We may be able to identify variants that were heretofore unknown. We may be able to identify variants that were not seen because they were below the level of the limited detection of the tests, and still there would be benefit in those patients.

You've heard all these questions, and I would

dare say we've also heard some people proclaim the answers with all of those same things, saying, yes, we can. Yes, we can identify mutations, and those mutations make a difference. And then the next question is, where's the published data to show?

Where do we need to go? Testing, testing, testing. It sounds like the old screen of the emergency response system. We need to understand and truly do the tests. We cannot make up the science on this. And what's amazing to me is heretofore, we have done a reasonably good job of working in silos to try to get these answers.

So we're still to the testing stage, and it's made it very difficult because we've seen that it's been difficult to get some of these tests to a point that the FDA or payers are willing to be able to say this is something we can trust.

The usual testing methods, define the tests; determine the patients; and we've talked about all morning long, define the intervention based on the tests, and then let's collect the clinical utility or let's show the outcomes. This is something that's

reasonably simple. We understand this.

Now, this is something -- and this is the way I look at next-generation sequencing testing. We spent all morning talking about this. We have the pre-analytic, and I simplified the pre-analytical into the hybrid capture amplification version, the variation. And once again, I'm not a molecular pathologist. I don't understand the ins and outs of it, but we heard from this morning that there are large differences in the way you prepare samples. And that doesn't even get in on how you sample the tissue itself.

Then you have the sequencing. I can do it on one instrument, or two instruments, or instrument X. There's different ways of doing the sequencing. I can look at targets. I can look at whole exome sequencing if I want. I can look at RNA. I can do a combination of above. I can do different panels, maybe only looking at 5, 15, 500, 1,000. What's my depth of read? What is my version and my variation?

And we all look at this and we say, any one of those variables in the sequencing could change the outcomes. Now, we heard this morning that if we're

looking at SNPs, we're probably pretty good, and I would agree with that. If we're looking for single nucleotide polymorphisms in somatic tumors, we probably have reasonably good concordance depending on what instrument and what methodology you can identify it.

But the problem is that a lot of the mutations that we have proclaimed we haven't found in the past and companion diagnostics missed are mutations that are a little bit different such as insertions and deletions, duplications. And we've heard -- and there were several papers that had been published to identify the concordance in instrumentation independent of the instrument itself have concordance rates that are, at best, 20 percent. And the false positive rate on these are well over 50 percent.

Then you have the whole informatics pipeline where you can look and you can run it with algorithm 1 to algorithm X, and you can identify different variant calls that may or may not be duplicated if you were to run a different bioinformatics pipeline.

So as I think about each test, I will look at it as something that looks like a very complex

chemistry equation. We have the pre-analytical that's done by hybrid capture version 1.6. We have sequencing that was done on an Illumina, or a Thermo Fisher, or Genomics Health, or whatever else boxes are being made out there. And it was whole exome sequencing version 2.2, and it was using their fourth panel or iteration. And the informatics pipeline was made by company A, B, C, and it's version 1.2.

Then you could have a completely different test that has a completely different one done by amplification, made by a different company, using targeted exomes, looking at a different panel and using a different informatics pipeline.

Until we understand that -- and I think we do understand this, that we really, really can't aggregate clinical utility data unless we define the testing. So if a manufacturer comes in and says, I really, really want to show that my biomarker panel can do those things that we said it can do in the Swiss army knife, then we have to know how that exact test with the exact version leads to the exact outcome. And it becomes very difficult because we're trying -- particularly in

all these other databases, we're aggregating genomic data that we do not know if it is something that we can aggregate because we don't know what the testing's like.

This slide is something that I think we all understand, and I think it's very important to discuss. And we haven't spent a lot of time talking about this. There's a blue line in the middle of this rectangular box that I call the clinical threshold for benefit. This is not the limit of detection of the test. It just basically says that anyone above that line, if they have a biomarker and we treat them with a treatment, will or has a high likelihood of responding to it. And anything below that line, even if it is existent, even if it is real, will not benefit from the therapy.

So the idea is that increased sensitivity may not lead to better outcomes. Now, we know that there's not a complete overlap between the next-generation sequencing and the companion diagnostic, and we know that the companion diagnostics are not perfect anyway. And we can argue that companion diagnostics may have

higher sensitivity, or lower sensitivity and higher specificity, or maybe not. Once again, as far as I know, we have not seen any dramatic comparisons between the two in such a way that we can determine not concordance between the testing but concordance in the outcomes based on that testing.

We may be picking up different types of alterations that are not targetable. We may be looking at lower allele frequencies. We may be even looking at simple biopsy differences from a heterogeneous tumor. We may not see better overall outcomes. And therefore, the way I look at this in my very simplistic way is, what matters most is the patient and whether or not the patient will respond to therapy. In other words, we're asking binary decisions, binary questions: if biomarker, then treatment, and we hope that that "then" leads to outcome.

So do we need trial endpoints? Should they be analytical or clinical? Can we simplify anything? Truly, as we start looking at the best studies, we know that they are randomized controlled trials that compare one arm to another. We can look at biomarker

comparisons of one testing to another, but we need to do it two ways. We can say maybe my test is as sensitive. Maybe it's too sensitive. So we can do that. We can look at retrospective analysis.

There's something that we're seeing right now, which I don't know if we can do, which is what I'm going to call the general consensus of a standard of care without published data. In other words, this is what I call the digoxin era. We just all agree it works, and it takes us a hundred years to show that it does make a difference.

So my last slide, because I'm over, do we need trials? Absolutely. And I don't know how to get this really shown that it will benefit our patients unless we do trials. And I would dare say that the endpoints of the trials should be clinically based.

Companion diagnostics, when they were shown to be beneficial to patients, had a very important clinical outcome that was associated with those trials. And if we're going to replace the companion diagnostic, we ought to probably also look clinically to make sure we can do it.

Can we simplify the standard of care without published? And the answer to that is maybe. We can find a way of simplifying testing. If manufacturers can find a way of saying we can work together rather than apart, and we can find ways of being able to aggregate data sets, it's going to be a lot easier when we put them together in combined data sets to show that there is clinical utility. And we can really figure out this pesky sensitivity, specificity issue that is such a big deal.

The last thing I'll say is it would be nice if we could get to an era where we're collecting data on all patients. Big data can be very messy, and big data can be incredibly difficult to understand. But bottom line is if we can collect a lot of information and standardize it as much as we can, we may possibly get to the point that we can work together to answer these important questions. Thank you.

(Applause.)

Panel 3 Discussion and Questions

DR. TZOU: Thank you, Dane.

So before proceed to the questions, I'll ask

the panelists to introduce themselves, starting with Dr. Sklar.

DR. SKLAR: My name is Jeff Sklar. I'm a professor at Yale. And I've been in molecular diagnostics for many decades, and I founded the tumor profiling lab at Yale about six years ago. It's an offshoot of molecular diagnostics. And I also run a lab that's heavily involved in doing companion NGS for clinical trials, including the NCI-MATCH trial.

DR. TSIMBERIDOU: I am Lia Tsimberidou. I am a hematologist/oncologist and tenured associate professor at the MD Anderson Cancer Center in the Department of Investigational Cancer Therapeutics. In 2007, I initiated the Personalized Medicine Program in the phase 1, in this department, by initiating the first IMPACT trial, Initiative for Molecular Profiling in Advanced Cancer Therapy, where the goal is to do molecular profiling before we selected the phase 1 clinical trial for our patients. We demonstrated that using molecular profiling, we had encouraging results in terms of response, progressive-free survival, and survival.

Later on, I started the IMPACT 2 trial, which is a randomized trial evaluating molecular profiling in patients with metastatic cancer. This trial is conducted at MD Anderson, and it is sponsored by Foundation Medicine. And I treat many patients with advanced cancer using the personalized medicine approach.

DR. BLUMENTHAL: Hi. My name is Gideon Blumenthal. I'm in the sister agency of CDRH. I'm in the Center for Drug Evaluation and Research. I'm a hematologist/oncologist. I focus on thoracic oncology. I'm a clinical team leader for thoracic and head and neck cancer, so we oversee the drug development programs in thoracic and head and neck cancers. I also, as part of my professional development, see patients at the National Cancer Institute in the Thoracic Oncology Clinic.

MS. KREUZ: I'm Greta Kreuz, and I'm not a doctor. And a lot of this stuff is way over my pay grade. But I am a patient advocate, and this is my first time to do this, so bear with me.

I'm a stage 4 lung cancer, non-small cell

lung cancer survivor. I've never smoked, just to give you -- very brief. Never smoked. I was a runner and in the news business. And in 2012, I went in for a regular physical and had no symptoms, but they found a tumor. And I had a lobectomy and was stage 1B, so I didn't require any follow-up, except monitoring.

A year and a half later, I went back to work, and I did a series and won an Emmy for it. But it didn't really impact my life too much. And then a year and a half later, in the fall of 2013, I started to have symptoms. I had a cough, and I had difficulty climbing and hiking, and wasted several months. The scans were all clean, so it was not a pretty time. And finally, New Year's Eve of 2013, I was diagnosed with stage 3. They found a tumor in my airways that had gotten into the lymph nodes.

Anyway, they attempted surgery in May of 2014, but when they opened me up, they discovered that it had spread into the pleura, the lining of the lungs, so they just basically closed me up, and that began this whole journey of information overload and am I going to die tomorrow kind of thing.

So I underwent chemo basically. I had genetic testing done. I have the BRAF mutation. But I started chemo, and I have been on it ever since. I went to a maintenance regimen in October of 2014 of Alimta, a single drug every three weeks infusion. And I've been on it ever since. I'm taking a break now, a brief break, just to try because you don't know what it does to your body after all these years. I retired from Channel 7 in part because of my health and in part because I couldn't tweet, or Instagram, or any of that stuff.

Anyway, so now I'm busy trying to do a number of things, but one of them is get involved in things like this because I'm connected with a lot of -- particularly lung cancer organizations, and I have done videos and so forth. I posted something on Facebook because I wanted to hear back from other people who knew I was going to do this. And basically, the bottom line is, people with lung cancer, and with all kinds of cancers I think, were in awe of what you do.

I'm amazed, and we're so grateful that so

much is happening now. My sister died of lung cancer when she was 49 in 2004, and she never had a chance. And so much has happened since then because of people like you. So I just wanted to thank you so much.

But there is an era of mixed emotions from lung cancer survivors, of desperation, and impatience, and confusion, and anger, and what's the hold up, and hurry up. So it's not a monolithic group of people out there. Just like you're having discussions and debates, so are the patients.

I guess I just wanted to thank you for letting me be here. I'm going to try to weigh in as best I can on some of this arcane stuff. But just thank you for letting me be a part of this.

DR. TZOU: Thank you for all the panelists.

(Applause.)

DR. TZOU: So we'll proceed to some of the panel discussion questions. The first topic concerns follow-on companion diagnostics. One of the issues raised by Dr. Dickson's talk was how different tests may compare, whether they may select different patient populations. The rationale behind FDA evaluating and

approving companion diagnostics is that they select a particular patient population for which treatment, safety, and efficacy have been established.

So for follow-on companion diagnostics, for example in just panels that may claim particular variants and say that patient's identified as positive for those variants might also be candidates for those corresponding FDA-approved therapeutic indications, the topics of discussion would be what sort of performance measures, what sort of agreement, would provide evidence of safety and effectiveness for a follow-on companion diagnostic claim, and what types of study design considerations would be important as far as where would one obtain the appropriate clinical sample sources, recognizing that, ideally, one would take the original clinical trial samples, but they may not be available.

So in the event that the original clinical trial specimens are not available, what are the considerations that would be important to consider whether or not outside specimens may be procured, whether they are reflecting the appropriate patient

population, whether they're important characteristics and changes in specimens to determine whether those are appropriate samples?

So I'll open it up to the panel for whoever would like to start off.

DR. SKLAR: Well, I think in terms of -- if I understand the question correctly, I think the -- I don't think this is a huge problem. I think if you're talking about just the validity of detecting a variant, as we heard in the discussion this morning, DNA is pretty much DNA. RNA is pretty much RNA.

I think probably the biggest consideration is the pre-analytic issue of what kind of sample you're using. If you're using -- of the original study was done on FFPE, it's probably appropriate that the follow-on uses FFPE. Any kind of qualification that's been detected in the original study like particular tissue that's problematic, interfering substances that are mentioned sometimes such as melanin, hemoglobin, those same kinds of warnings should be heeded when considering a follow-on test.

I think it's important probably to note

that -- and this is maybe slightly tangential to the subject, but when studies are done, I think it's important -- and trials are done, it's important that the analysis of the tissue, the way the tissue is handled and the way it's analyzed, should take into account how it's likely to be done in the field. When the information gleaned from the trial is actually utilized in laboratories throughout the country, that the trial is designed in such a way that the laboratories will be using the same technology.

I think that along those lines, it should be encouraged that the people designing these trials should consult with the testing laboratories and the pathologists to ensure that they're actually using the right type of material, the type of material that's going to be used later on. That's not always done. Some of these studies are designed with not that much interaction or consultation with the actual laboratorians. So I think that's an area we might be able to improve.

DR. DICKSON: I'm going to ask Jeff a couple of questions. So, Jeff, let's say I go through and I

compare the companion diagnostic in a patient, and all of a sudden I say I'm going to replace the companion diagnostic with something completely a new technology. And I get an insertion in the NGS panel that's made by company X. Can I trust that insertion to be real?

DR. SKLAR: Well, I think you have to do some kind of orthogonal testing if you get a discordant result like that. I think that -- another thing that I've thought about, another recommendation perhaps is that companies who sell these tests provide control material as part of -- if it's a kit, as part of the kit.

This is done for certain kinds of molecular tests like FISH. Slides will be sent that you can stain with a probe and determine that you detect the actual signal that should be there. Maybe the same kind of thing should be done so that you're actually -- a laboratory can confirm that you're finding what should be found.

It should be noted, I think -- actually, there are two levels of validation in practice. And that is -- I think we should keep this in mind, and

maybe it was implicit in everything that was said this morning, that there's the company who will validate their test. And then there's the laboratory that validates the test. And we're under CAP regulations, and we're scrutinized, and we have standards.

So anything the company does will be revalidated by the laboratory. And essentially, I think some of the gist of what this conference is about is how does a laboratory decide which technology -- which platform to purchase based on what that company has done in their validation. Then we will revalidate it. We'll confirm that we can find what they find. And until that's done, I don't think any laboratory would actually perform testing.

DR. TZOU: So perhaps I'll just provide some context to the range of scope for this question. It's certainly possible that in an FDA-approved NGS oncopanel, there could be a range of laboratories that might be interested in using it. I think there is one range that could be a laboratory that has more experience and expertise and might want to do additional internal validation. So that's one

possibility.

I think another possibility is that there could be a range of laboratories that previously had not had experience in developing NGS panels on their own and might rely very heavily on what the manufacturer said as far as the performance of this product. So it could be that NGS oncopanel was approved by FDA, that a lab that previously had not been involved in this testing might decide, therefore, that since this has been evaluated, FDA, they would just do verification. They would not do more extensive validation than what the manufacturer provided.

So in that scope of possibility, is there a difference in perspective as far as if -- for that type of use setting, for a lab that has not previously been doing NGS testing, an FDA-approved test becomes available, and they decide therefore I'm going to use it. And since the manufacturer said that the variant identification is as good or comparable to a companion diagnostic, and they would be relying on that, what sort of studies would be sufficient to support that context?

DR. KULKARNI: So apart from what you just said, that the lab may not have enough prior NGS experience, there's one more other additional consideration we should all talk about and discuss, is that the requirements by organizations like CAP and New York State are different if the test is FDA approved versus a laboratory developed test. That is an important consideration. The standards are a little bit less rigorous for an FDA-approved probe.

So that's something which I wanted to throw out here for discussions. In addition to that, there are several other things. For example, one example comes to my mind, which we deal with on a regular basis. For non-small cell lung cancer patients, our NGS panel has intronic baits where we could pick up ALK translocations, for example, by DNA sequencing.

We almost always reflex to the companion diagnostics FISH test to -- as an orthogonal method, not only because of orthogonal method, but also because some insurance companies require the companion diagnostic test for the drug approval. So we'll have to do it anyway.

Do I approve of that? Do I agree? I have mixed feelings. In our hands, because of the way we design it, we covered pretty much all the hot-spot breakpoints in the intronic region, so we are confident -- and because we have enough data sets, hundreds of patients where the translocation was found by NGS sequencing and had 100 percent concordance with FISH. But we cannot rely on all the vendors or all the other labs to go through the same rigorous, upfront validation.

So there has to be some standards where we can assure before we jump on to this expensive drug, which might not only be expensive, but might do harm to the patient. So those are our considerations, which we have to recommend.

MS. KREUZ: I just wanted to jump in here, too, because one of the things that I hear and I went through is I had genetic testing done by three different institutions. And again, when you're battling lung cancer and it hits you over the head, I mean, you're just traumatized, and you don't know what you're doing.

You're in there with your doctor, and you're just like whatever it takes kind of thing. And he says I think we should ship off your tissue to this place and have a genetic -- okay, fine. And you're not asking about insurance. You're not asking if it's covered. You don't know.

So as it turned out, it wasn't covered. But then I went for another consult, two other consults, and they both said, well, we're going to send your tissue off to X, Y and Z, or we do it here, or whatever; different panel, different number of mutations. And in retrospect, I'm wondering why must I have multiple genetic testing at different places. Weren't the first guy's responsible or reliable? And is it just a way for people to make money?

I just don't know how that works. I mean, what do you tell a patient? Why can't company A be the same as B and C, or why can't I rely on them, and why do I have to go through multiple genetic testing?

DR. SKLAR: Well, usually we lose money, as a matter of fact. But did you get the same answer?

MS. KREUZ: Well, I did on the BRAF, but some

of them again tested for more mutations. So some of them are more thorough or more extensive. So I don't know the whole answer to that.

DR. SKLAR: Well, that is a problem. It's a well known problem in the field. Different platforms will give different results. Different informatics pipelines give different results. There is a core of common findings, but there are often other findings. And I think to a patient that can be extremely confusing. It's confusing enough to us.

MS. KREUZ: And one footnote to that, too. It's been brought to my attention that one company in particular -- and I don't know how extensive this is. But when they released their findings, with that comes a recommendation that because they have, A, that you should do this drug; we recommend this drug. Even if maybe that drug isn't effective or other studies show it doesn't really work. I don't understand that, and what guarantees are there that that company knows what it's talking about?

DR. TSIMBERIDOU: From the clinician's perspective, I think it is very important to use the

most accurate and extensive molecular profiling that is available. And certainly, next-gen sequencing provides extensive information that a single companion diagnostic test cannot provide.

However, I think it is important to carefully collect this data and analyze to make sure that there is clinical relevance. And here is the value of prospective clinical trials to understand the value of certain molecular alterations and the interaction with certain drugs that are considered targeted. And I agree with you that simply because a certain company recommends that the drug is known to inhibit the function of the altered [indiscernible] gene, it does not necessarily mean that if we treat the patient with this drug, the patient will necessarily respond to it.

There are different levels of evidence. For instance, some people or some of these companies include even data from in vitro models or from the package insert where data are not validated in trials in humans. And I think we need to be very careful when we assign a drug to a certain molecular alteration regarding the clinical validity and significance.

That's why I think it is very important to prospectively collect data. I agree with the previous speakers that we need to standardize next-gen sequencing and establish certain rules that all companies should follow before we use them to make treatment decisions.

DR. SKLAR: I'd like to add one other comment. This relates to the validity of a single company's product, but also follow-on drugs, and that is, a recommendation that there be greater transparency in the validation process performed by these companies. I suppose with FDA oversight, that's likely to happen. But heretofore at least, many of these companies are not very clear about what's been validated, largely I think because they're hiding behind the fig leaf of an RUO test. And because they have to maintain the appearance of not providing a clinical test, they really are not very open about how the test has been validated.

I think it would be much more helpful if we knew exactly what the informatic pipeline was that they used, how they made calls that they made, what kind of

samples they used for validation. If they use cell lines, and these things are readily available, then you can go and get those cell lines and compare your results to their results.

So I hope that through part of this process perhaps, this larger process, that maybe they'll be more openness about how these tests have been validated.

DR. TZOU: Gideon?

DR. BLUMENTHAL: Let me put Dane on the hot seat for a second here. On one of your slides, you said we need prospective randomized controlled studies, gold standard of overall survival. I totally agree with that. Being here at the FDA, that's like Mom and apple pie. But at the same time, we also have to be practical. We do live in the real world, and it may be hard to design or adequately power studies to say I'm going to compare a cohort of patients treated with companion diagnostic X versus novel NGS assay.

So how do we -- with this struggle between what would be perfect and what would be good enough, how do we get around those types of issues?

DR. DICKSON: So my last point on that slide was although randomized controlled trials were the gold standard, there's the digoxin era. We use it because it works. And I think ultimately, we all look at each other and we say, next-generation sequencing is powerful. You had one speaker who said don't get me wrong, this is good stuff, basically.

So what I see happening -- and this goes back to Greta's point -- is I see that we've got to do several things. One, we've got to collect a lot of data. But in order to collect the data, we've got to standardize the tests. What we've got to do is we've got to, in an iterative process, understand. We've got to understand such things as what does the variant calling software identify? How well does it respond to a certain companion or to a targeted drug? What is the limit of clinical detection? Not limit of instrument detection, but when are we going to do it?

So I think the best way to do that is through registries and establishing registries are being done. But one of the problems I see with the registries that are taking place is that there has not heretofore been

a great deal of standardization of those registries. There's been a, trust me, my lab is doing it, and I'll trust you that your lab is doing it even though they really haven't done anything other until the NCI-MATCH trial, which Jeff's group has done.

I think the NCI-MATCH trial did an amazing job of saying let's go through and let's standardize that. And I think going towards something like they did at the MATCH trial -- but we may have to choose three standards. We may say hot-spot testing, and then whole exome sequencing, and a more comprehensive or exomic somatic gene analysis. I don't know what those three standards are, but I think if we can get to three standards we all agree to, then we collect that data. Then we move forward without putting patients in harm's way. But we've got to have that standard and we've got to have that data sharing.

DR. TZOU: All right. So I'm going to move on to the next question. The first topic was in regard to follow-on companion diagnostics claims. We know NGS panels may also include other genes and variants for which the evidence may not be as well established. So

the general question is, what does the panel think as far as the level of evidence or rationale to support inclusion of those variants.

So one aspect would be does the comparability of its analytical performance with these other non-companion diagnostic variants, is that a criterion for consideration? And also, would something, for example, like the gene and variant being included as part of the criteria for a clinical trial, would that be an example rationale to support a variant?

So again, in this context, the scope of labs that may be entertaining using an FDA-approved panel, they may not have the sort of tumor panel expertise, may not have the type of in-house curation staff, which in the previous panel we heard may not be scalable anyway to support a more nuanced interpretation of some of these considerations.

DR. DICKSON: One of the things I think that we'd all like to see is a pan-companion diagnostic tool, something where we can say this test, because it's been standardized well enough, we can use this test to then act as the inclusion criteria in a

clinical trial for which then we automatically have the information that we need.

But I think that comes back down to how well can we standardize the test, how well can we have concordance across labs, particularly when we look at labs that may have a great deal of experience, like Jeff's lab or Wash U is doing, and then maybe a smaller lab that may not have the same capabilities.

So I think it comes back down to say can we standardize the test well enough that we can then collect data in such a way that we can be sure that that test does lead to an outcome that's beneficial to patients.

DR. TSIMBERIDOU: I think, again, from the clinician's perspective that it's very important to determine the variant. For instance, the BRAF drugs works best in patients with BRAF V600E alteration, including in other tumor types that I have personally treated. But if you try to treat other alterations, you are not sure of the BRAF gene. You're not sure that you will get the same results.

So I think when you design a clinical trial,

it is important to determine what are your expectations regarding the alteration of specific variants. I think we have learned this only from clinical trials, prospective clinical trials. Because even the phase 3 randomized trials are very challenging, and we're all aware of a study that was already published from France, that demonstrated no benefit when you use targeted therapy compared to the other arm. However, the selection of the drugs were suboptimal.

So we cannot generalize those results. What is important is to build the N of 1 databases. And ASCO has the ASCO cancer link, that they're building a database, and hopefully we will learn a lot from sharing the data. MD Anderson is building their program. I think as a community, we need to create a database and share those N of 1 data and use them to treat the next patient who comes better.

I do it in my practice, looking at the 3700 patients we have previously treated with molecular profiling. And I'm trying to identify how many patients have a similar alteration, a similar tumor type, and treat them with the best treatment possible.

But we need to be able to create these databases and share these data in my opinion.

DR. BLUMENTHAL: So I think the question was centered around, let's say you have a companion diagnostic claim for like exon 19 deletion, and can you extrapolate, can you borrow that data for other indels in other genes to get that sort of tier 2 claim. It would be interesting to hear thoughts on that, being able to borrow because you can't validate every single variant in every single gene. And that makes a lot of sense.

The second bullet about the clinical trial NCT number, would that be sufficient, again, you would like to know within a clinical trial which variants. If I'm doing a MET study with a MET inhibitor, you might test exon 14 deletions. You might test copy number variants. But unfortunately, in clinicaltrials.gov, there's not that type of granularity. So I think referring to the NCT number may be sufficient, but would be interested to hear others.

DR. KULKARNI: I wanted to add three quick

points. Databases like ClinGen, which would be NCBI ClinVar database. Outcome data would be the first step, and we have commitment from 50 large clinical labs, academic and commercial, to have that access. That's point one.

Then there are -- as you mentioned, we need more N of 1 reports out there. Just a way of advertisement here, I'm the editor-in-chief of a journal called Cancer Genetics, and we have just begun to start accepting these N of 1 case reports, which are very standardized, not labor intensive. So anybody who has all these N of 1 case reports, please send our way. We promise rapid revision and acceptance, so that we can have more of these out there.

There are additional efforts ongoing like that. There's a whole journal dedicated to this called Molecular Case Reports. Elaine Mardis at Washington University is the chief editor there. And that's the same concept, that they would accept all these N of 1 case reports because there's a huge need for these reports.

We in our lab have maybe 30 such N of 1

anecdotal reports where we have demonstrated exactly this, where gene not variant based therapy has worked. Then also, in other cancer types, we have a patient with thymic carcinoma who has KIT deletions. So imatinib which works, and just worked perfectly fine and was a miracle. This is a rapidly growing mediastinal tumor, and the patient who was in a hospice in two months went home and started her regular life.

So we have those success stories sitting in our lab. We just are so pressed with our time, we just don't get enough time to publish those. So with these kind of efforts and with collective efforts from everybody, I think we will build up that evidence.

DR. SKLAR: Just to add my perspective on this, I think it's terribly important -- I had a conversation at AMP -- I think it was on exactly that case of the -- I think it was a thymic carcinoma that was treated at Washington University. I think it was a BRAF mutation actually.

DR. KULKARNI: No.

DR. SKLAR: It was KIT.

DR. KULKARNI: KIT.

DR. SKLAR: So I think one of the reasons why this is so important, the MATCH trial may do this in a systematic way, but these anecdotal reports of treatment of idiosyncratic mutations that appear in tumors which are not usually associated with, it's very important because it's the rationale for broad-based, next-generation sequencing of tumors.

That's why you want to do it, to find that mutation that you don't expect. It's easy to look for the BRAF mutation in melanoma, but to look for BRAF in a thymic -- well, it wasn't thymic carcinoma. But to look for BRAF in a brain tumor or something like that, and then to know that it's targetable, that's extremely important. That's why you have to test for all --

DR. KULKARNI: Yes. And sometimes you might find out that it doesn't work. That's also important information. So you might know better than me, BRAF V600E in hairy cell leukemia does not respond the same way to -- whatever the drug is.

DR. TSIMBERIDOU: I think you bring up an important issue, I think, and here is the value of next-gen sequencing, looking at the other alterations

and perhaps explaining why a patient with a certain alteration may not respond to the targeted agent. For instance, if patients have at the same time a P53 mutation, then they may lose their response.

I had a patient with salivary lung cancer with a BRAF V600E mutation who responded very nicely to vemurafenib. I treated him on the BASKET trial by Genentech, and his response lasted for about two years. And I repeated the molecular -- that was the single alteration he had in his molecular profiling. When I retested his tumor, he had an acquired P53 mutation.

Now, was this the only reason? Well, I think we need to learn a lot about tumor biology. There are other data evolving about the role of functional genomics and proteins, certainly the gene and genetic analysis alone is not the answer. I think we need to understand the dynamics of tumor biology, the emerging tumor biology, as we treat these patients because the profile changes.

All the other challenges we have, for instance, performing a biopsy of several areas of metastatic disease versus the original tumor and

understanding the differences and understanding -- I think all healthcare providers and patients, we need to understand those challenges, that a single molecular alteration from a single tumor is not perhaps adequate.

Ideally, we should biopsy many metastatic sites, and that's where the liquid biopsies come along, which is not the topic of this session. However, I think we need to also do better -- studies with a tumor to make sure that we would be able to make meaningful treatment decisions.

MS. KREUZ: If I could just jump in on that, and I think I'm understanding most of this. But what I hear a lot of from my colleagues is a lot of them are just trying all kinds of things. They have a mutation that works with one type of drug, and now they're trying it for their -- their crossing over. They're trying different things.

I can't even cite examples because I get them all mixed up. But they're not locked into -- you talked about the melanoma and the lung cancer, the companions; it seemed to be some overlap. And I think there's a lot of that in cancer, and a lot of the

people who have lung cancer are now looking at drugs that are used for other cancers and trying them for themselves. And I guess it's working in some cases.

So I think that's an exciting -- I guess what I'm saying is that's an exciting arena for patients, and we hope that scientists and doctors really, really push and explore that. And I understand there are so many variables here, but that's really an exciting area for people with cancer, that may not be locked into their own little drugs and their own little kind of mutations for their particular type of cancer.

DR. TZOU: Okay. So I'm going to ask if understanding that there's still evidence being developed and that the emerging evidence base may change, if it were to -- a new therapeutic indication were to be approved in the future or if more evidence were to support a potential new companion diagnostic indication, does the panel think that additional evidence might be appropriate for a NGS panel to support a more explicit new emerging companion diagnostic indication?

DR. KULKARNI: I'll take a stab at the first

one. I think we learned this morning about the matrix effect on tissue types. I would say you may not need the same level of extensive validation, but you at least need some version of that where you can make me as a laboratory comfortable that the matrices don't interfere.

DR. DICKSON: I would say that in an iterative process not only do you have to worry about the matrices, but also if we're to have testing that have different sensitivities, trying to identify when and where -- I mean, this is the HER2 argument. We spent a lot of time recognizing that it was only the -- we know that the HER2 -- 1-plus may respond, but it's the 3-plus that really do respond to therapy.

So there is some belief that, yes, allele frequency is going to be important. So I would say even after an initial validation, even if it's laboratory, there still has to be a clinical validation that comes along to in an iterative fashion identify what is the sensitivity level that is important so that we don't treat patients with drugs that are ineffective, that we don't waste their time.

Especially in lung cancer patients that are so important, often times they lose their performance status so quickly that we don't want to mess around with treatments that are not effective, particularly with targeted agents who have very little likelihood of benefitting if they don't have the alteration.

DR. TZOU: Since Dane just brought up this issue as far as differences and potential sensitivity, I think what I heard from Shashi is that if there is emerging evidence, it may not be as extensive as the additional validation, but you may want to do some -- not as extensive but still some confirmation that it works in the potential new companion diagnostic indication.

But Dane brought up a concern that while maybe you may be detecting different levels of sensitivity, you might be identifying different levels of patients. So do those two -- does the first one depends on you can confirm it, but, Dane, if there's still a concern that you may be identifying a different proportion of patients, that the first one might not be enough, or is this -- how does this issue of potential

levels, detection levels and different proportions of patients factor into people's consideration?

DR. BLUMENTHAL: Yes, I think that's a huge issue. I mean, if you're ramping up your sensitivity a hundredfold, that could potentially expose a lot of patients to excessive toxicity and no benefits. That's a great concern.

DR. SKLAR: I think this relates to something I noticed in the paper that was posted by the FDA, and that is -- maybe this was a misprint, but it said, "Refer to NGS testing as qualitative in vitro assays." They're not qualitative, they're quantitative. I think that's a very important aspect of these tests. I think sensitivity and a level of limit of detection is very important.

So a sensitivity that's a hundredfold greater so that you can detect something that is 0.01 percent VAF in the sample would not be a positive result. It's very important to have these thresholds for purpose.

Now, what they are I think is an extremely interesting question. And that has to be evaluated, and that also affects whether a test is sufficiently

sensitive. This is probably overly sensitive. But I'm reminded of the example of T790M present in subclonal amounts in lung cancers that are resistant to anti-EGFR therapy and 10 percent VAF frequency within the tumor tissue, yet the tumor is completely resistant, and it's growing.

So it's very important to be able to detect things at that level to determine if you have -- in that case is a resistance marker. But if you have a marker that predicts sensitivity, the frequency is important and above which a drug is likely to work. So I think this is a very important issue of sensitivity. These are quantitative tests, and not even to mention CNVs and what that means.

But I think that something that's overly sensitive, if one treated this like a qualitative test and said that's a positive test because I've gotten a signal, albeit at a very low level, that would be inappropriate. So we do have to set thresholds.

DR. TZOU: The level of signal was important.

DR. TSIMBERIDOU: And I think here is the value of next-gene sequencing because you are not

looking only at one companion diagnostic test, but you look at the broad spectrum of molecular alterations. And clinically, there are published data showing that the significant subclone, for instance, for patients with melanoma, is what leads to patients' death at the end after multiple responses and disease relapses.

So I think we need to know, even the subclone, even at the higher sensitivity, what is there. But we need to focus on what is the driver. And here again is the role of bioinformatics and functional genomics. Because one gene alone or one alteration alone does not determine the biology. You will have to put things in perspective and look at the global biology, and take into consideration that this is changing as we treat these patients with targeted therapies. It is not stable. It is very dynamic.

DR. DICKSON: Lia, I think you bring up an excellent point that we haven't talked about yet, which is the whole reporting mechanism because 80 percent of the oncology in this nation are practiced in places that probably don't have molecular tumor boards. They don't have the ability to understand a report.

So what scares me a little bit is if it's clearly on the report that there's an EGFR mutation at a 5 percent allele frequency, will a provider understand that that is something that some people would have called negative? And do they understand what to do if they have a BRAF and EFGR or it's a VUS, so a EGFR mutation?

I don't think that our clinicians are in any -- aren't able to interpret complex molecular reports, so we need to really I think also look at -- as we're looking at sensitivities and looking at iterative approaches for this also look at reporting to make sure that our reports are complicated enough, but not so complicated that we hurt patients.

DR. TSIMBERIDOU: I agree.

DR. TZOU: I'll thank Dane for again prompting another question topic. You just brought up this question of reporting, again, there is variability as far as how much tumor board or expertise, as far as crafting reports. There is variability across different NGS panels; and as Greta has experienced why she's getting tested multiple times and what do those

different panels mean.

So what are the panelists' thoughts as far as if there are different panels, they cover different things. They may be more or less comprehensive. How would users, whether they're clinicians or patients, be able to recognize these differences? And even within a particular panel, if particular genes, exons are not covered adequately in that particular run, how that might impact assay interpretation.

Depending on the level of evidence, whether it's companion diagnostic or not risen to a companion diagnostic, how people would be able to understand those differences; and whether they are de novo variants that have not been previously reported as opposed to defined hot spots, all these, does the panel think some of these are more or less important to key in so that the report users might be able to recognize some of these key differences?

MS. KREUZ: I just wanted to touch on that because that is something that's been brought to my attention from cancer patients, is who should be considered an expert when they get a report on genetic

mutations? How can the patient be assured that their provider knows how to read that and can interpret it accurately? I mean, not everybody is an expert in specific mutations I would guess.

If you have an oncologist who's more of a generalist, I mean, are they competent to interpret that and say, oh, this means this, so now you should do this? Or are there certain parameters or standards that are set to make sure whoever reads that knows what they're looking at?

DR. TSIMBERIDOU: I think this is a very critical point. In our institution, for instance, we have tumor boards where we present the results. We have annotation over the alterations, and we discuss what is the clinical value and significance of an alteration and what would be the best treatment. And in many other academic institutions, the same thing happens. There are tumor boards.

Also, ASCO's upcoming double blind trial with principal investigator Dr. Silski is going to address this issue, as patients be enrolled from several -- even private practices or collaborating

institutions. And there will be a tumor board that will be predetermining or setting the rules about matching molecular alterations with study drugs.

So there are efforts like this, but I agree with you that this is not available to all patients and all physicians. And that's where perhaps tumor boards are very important.

MS. KREUZ: And being not a medical person here, do tumor boards always agree? Or if you go to one place, they're going to say one thing, and you go to another place --

(Laughter.)

DR. TSIMBERIDOU: Well, we can find out.

DR. BLUMENTHAL: As someone who spends a lot of time on drug labels, I can tell you that we often get -- there are many federal dollars spent on making sure the drug labels are truthful and accurate and deter, for example, untruthful promotion. It's very tricky, and people ask, who is the label -- what's the drug label for? Is it for the patient? Is it for the prescriber? Is it a legal document? Is it for the payers? It's an interesting question. So it's very

tricky. And I imagine the device label is even less widely read than the drug label.

(Laughter.)

DR. BLUMENTHAL: Of course, you could put black boxes, you could put warnings till the cows come home, but I'm not sure how frequently this penetrates the psyche of the end users.

DR. TSIMBERIDOU: I think it's always for the patients. That's all we do.

DR. SKLAR: So a little bit of this speaks to the issue of transparency and validation by the company. We want to know that they test what are generally, for instance, as troublesome areas: homopolymer repeats, for instance, GC-rich sequences, things like that. And if they tell us whether they've done that -- even if we -- if they don't tell us, if it's a conspicuous absence, then we can query them about that type of thing. I do think they have to have disclaimers about this. If there's a known hot spot that's actionable and they can't detect it, I think they have to tell you that.

The related topic that kind of was touched on

is reporting. This is a message to the laboratory people, how you write reports. I think in this new era, we have to be a little bit more explicit about what these things mean. And without necessarily directing therapy, we have to explain the results.

I can tell you that there's great pressure from hospitals and others that we turn over these reports very quickly. It takes a long time to produce these kinds of things where there are so many variants. The report can get very long. They want a short report. Even some oncologists want a short report.

But I think that we really do have to explain what's there so that it can be interpreted by somebody who's not necessarily an expert in this area. Certainly, if you're providing information to outside institutions, not your own, where you don't have access to a tumor board, I think we have to write better reports.

DR. TZOU: So if I could follow up on that, if there are different types of users and labeling may be more concise versus more complete, does the panel have priorities as far as what would be the most

important highlights to feature in a more concise report to someone who is not as technically versed versus someone who may be more? So on this spectrum, are there high points to have in concise labeling, concise reporting, as opposed to more comprehensive approaches?

MS. KREUZ: Can I just ask -- because I know nothing about labeling. But whoever writes the labels, are they also tasked with marketing it, or it's just a strictly scientific bam, bam, bam? Or is it supposed to be phrased or worded in a way that would be appealing to whoever might buy it?

DR. TZOU: So there's labeling and there's promotional marketing activity more broadly defined, which includes labeling, but perhaps other activities as well. So for the package insert or device labeling, it may be more. Some versions may be more technical, but it may be possible that there could be, if the panel thought appropriate -- does the panel think it's appropriate that there may be patient-oriented labeling or clinician-oriented labeling?

So there could be more technical labeling for

operators or users. There might be information provided to patients or clinicians. So it's possible depending on what the panel thinks as far as who should be targeted or what types of information should be presented to whom.

MS. KREUZ: Yes. I mean, I would just caution as a patient or a user that they'd be cautioned to just try to keep it straight and not try to slant it. Because for credibility purposes, I think the general population deserves just keeping it straight. It doesn't have to be uber technical, but this is it, this is this, we don't know this, we do know that, without trying to sell it.

DR. BLUMENTHAL: At least in the Center for Drugs, we have a whole office for promotion, looking at overly promotional claims. We try to keep our labels as scientifically accurate and truthful as possible. There's intense scrutiny on companies if they oversell the truth.

DR. DICKSON: I think there's been -- I won't call it a disturbing trend, but I'll just call it a trend, which is the trend for identifying more

mutations, finding more targeted therapeutics. And I think most physicians have in their mind a binary decision-making capability, if, then. If mutation, then treat.

I think that we've got to be real cautious as we're looking at reports and looking at validation to make sure that we're not inadvertently trying to report something just to report something that a physician, if he or she misinterprets, would lead to an inappropriate use of a drug. I think we've got to get better at saying what is really a positive and what's really a negative, and what's something, and then how do we communicate that in such a way that the physician recognizes that many of these things are not as binary as they would like.

DR. KULKARNI: I completely agree. It's almost like an arms race right now in this space where my panel is bigger than yours, and my panel detects more targeted therapy than yours. So we have to be careful.

I see there are two separate aspects we're discussing here. One is guidance for the

recommendations, for FDA to talk to the company vendors, but also the second aspect of that is for the lab people to write reports. So I think those are two separate issues. For the lab reports, I think we are very cautious about what we recommend. We have one disclaimer, which is almost on all reports, is that these results should be used in conjunction with pathology clinical findings. Because we don't see the patient. We see the variant. The clinicians see the patient.

So even if a hundred times more sensitive test detects a variant and the patient is doing perfectly fine, then it doesn't mean anything. And it reminds me of -- I don't want to give up my age. But 20 years ago, we did a study, and I was in Imperial College at Hammersmith Hospital in London where we looked for BCR-ABL fusions in normal individuals, and we did find that.

DR. TSIMBERIDOU: Yes.

DR. KULKARNI: And it was ASH abstract -- ASH plenary session. I remember 20 years ago.

So you can find these driver mutations in

normal individuals with no clinical phenotype. So we have to be very, very careful. So we do our job. We write all these disclaimers, which are uniformly seen in our reports. But I think the vendors also have to use a lot of caution, especially when this is getting so complex.

DR. TSIMBERIDOU: I agree. The same for BCL2, found the normal people, not only patients with follicular lymphoma.

But I'd like to go back to Dane's comment. I think it's important to do more inclusive molecular testing. For instance, I have seen molecular testing for patients with advance cancer evolving since 2007 when I started this program, the phase 1 clinical trials program. Now for instance, we have patients with EGFR alterations, and we have access to these drugs through clinical trials.

I have a patient who did not respond to anything else, but now has responded to an EGFR inhibitor, has a response lasting for over 10 months. So for this patient, this makes a big difference, and we need to understand this. This is what personalized

medicine is about.

DR. SKLAR: I think there's also another argument in favor of larger panels. I'm not talking about gigantic panels and not just panels for self-aggrandizement. And that's the validation issue, which we've really been talking about. And that is that if you have a very small panel, and it only includes the variants that are completely actionable and targetable, and yet there are other genes and pathways that are likely to at some point turn up to have mutations that could benefit from the therapy, to not have those in the panel, the original panel, means you're going to have to revalidate.

So if you have a panel that's a little bit more inclusive in a judicious way, then I think you may avoid the constant revalidation, which is expensive and time consuming. So I think you want to be able to have kind of saltatory progress, you have a panel that's of a reasonably large enough size that you wanted to revalidate in six months, but maybe you will in two years. I think there is an argument to have a slightly larger panel.

MS. KREUZ: I would like -- oh, I'm sorry.

DR. KULKARNI: I don't debate that. That's what we have at Wash U. And based on the clinical indication to avoid a 300-page report, we use in silico masking of the data. And especially, also economics plays a big huge role. We all are losing money right now in this testing, and we have to be conscious about getting paid. So if you are within the 5 to 50 range, there is a chance that you might get paid. Right?

The benefit of being at Wash U is that most of our patients are also enrolled, are consented, so we do, based on the initial results, go back and take a peek, and we provide all the extra results to our patients anyway. Not everybody has the same luxury.

MS. KREUZ: I just wanted to -- my personal take on that is do it up front at the beginning and do it all; although I was talking to a colleague who took just the exact opposite. She said they wanted to do these very extensive genetic testing on her, and she wanted treatment now. And it was taking longer to get the results back when you do these very extensive tests. I don't know how true that is. Maybe they've

sped things up now. But she was like, "Why do they have to do all that stuff when most of it's not actionable and you're taking precious weeks?" And I need to get on treatment because I'm getting worse every day kind of thing.

So there is a balance to be had there. But I kind of think, like you were saying, they might not be actionable now, but at least you've got a profile, and down the road something comes, you can plug you right in.

DR. TSIMBERIDOU: I think it depends on the rate of progression.

MS. KREUZ: Right.

DR. TSIMBERIDOU: For instance, if a patient can wait, it usually takes two weeks from the time the tissue goes to a lab to get complete molecular profile and next-gen sequencing. If a patient has rapidly progressive disease, of course you treat them with induction therapy, and in the meanwhile run the test from the baseline biopsy, and optimize your treatment plan as soon as you get the response back.

DR. TZOU: So I'll open it up for one or two

questions from the audience. If there are any, please come up to the microphone, and please introduce yourself before the question.

MR. KONIGSHOFER: [Inaudible - off mic.] Yves Konigshofer, SeraCare Life Sciences. Starting with a drug that is using inhibitor that is for patients who are negative for mutations in KRAS, there's first this question what is KRAS? It's 100 kilobases of gene within which the actual companion diagnostics only look at several bases. Within those several bases, there are two companion diagnostics that look at those. They have different limits of detection depending on the mutation.

Now going to NGS, those limits of detection are somewhere between 1 and 5 percent for the companion diagnostic, but for an NGS-based test, it might be a certain number for all of these mutations. And I think typically if you have a companion diagnostic at 1 to 5, you want to at least be able to perform at 1. But if you perform at 1 and detect something, that you have 1 percent mutant of this one, by this companion diagnostic, you'd be negative. By the other one, you

might be positive.

But sometimes they might get the same result. Sometimes they may get a different result. And then there's this desire, as Jeff you were saying, with why did the FDA put in there this word "qualitative," because I mean, yeah, for the physician at the end that's getting the report, you want to know what to do, is it there or not there for this package insert.

But how do you match performance on a NGS-based test for something where the actual variant mattered for the companion diagnostic?

DR. DICKSON: One of the problems we've run into is we're comparing apples and oranges. We know the companion diagnostic hits and identifies a mutation, and we know if we treat with a drug, we can see some expected outcome. We know that with NGS if we identify a different mutation or a different alteration, and we have more sensitivity, many of those patients are going to respond.

Here's the question, which I don't think any of us know, which is, which is more important, sensitivity or specificity? Because if I treat a

patient who is truly a clinical false positive -- when I say a clinical false positive, I mean someone who tested positive, but they're not going to respond to drug for whatever reason.

If I put those patients on erotinib for two to three months, there's a high likelihood they're going to progress and progress reasonably quickly through those therapies, and there's a high number of people that may not get second-line therapy. So I think that's something we've got to figure out, which is what's the sensitivity and the specificity? Do we go for greater sensitivity and identify more mutations or do we go with greater specificity in saying we see more patients who respond.

I think that's a great question, and I don't think anyone knows the answer. I think patients would say I'd prefer more sensitivity than I would specificity because I want options, treatment options. I think as clinicians we want to say, okay, let's just collect and continue to collect the data to understand exactly what those EGFR mutations are being found in the panel by NGS, what is happening with them and can

we, as time goes on, better understand, now we've got those results; what do we do with them?

MR. KONIGSHOFER: Would the tests have to match the performance per mutation, if it were a follow-on companion diagnostic?

DR. DICKSON: I mean, if it was a complete match overlap, then you could say yes, it can replace it entirely. But we know that we're looking at things that are not complete overlaps. There are things that have greater sensitivity, they pick up different mutations.

Do we know that all those mutations or those alterations are therapeutically not actionable, but therapeutically treatable? They will respond to it. No. Well, a lot of them are. We don't know the sensitivity issue yet.

So those are some of the things that I think are important to really figure out as time goes on. If it was a complete overlap, then we've lost the power of the technology. So we've got to understand that we don't want a complete overlap, but we also don't want to get so sensitive that we lose response of drug.

DR. TZOU: All right. I'm going to make one administrative announcement. For those speakers who are participating in the public comment period, if during the break you could just check in with individuals here sitting by the computer, make sure your slides are ready. And for those who are also -- some presentations, we'll ask you to line -- be seated along this side of the wall, so you'll be ready to make the presentation.

So please join me in thanking the panel for all their insight and perspective.

(Applause.)

DR. TZOU: We'll be taking a break until 3:30.

(Whereupon, at 3:03 p.m., a recess was taken.)

Open Public Comment

DR. PATHAK: I hope you all had a good break. The sessions from this morning have been very interesting: pre-analytical, analytical, and clinical. It is time, however, to commence the public comment section of this workshop.

My name is Anand Pathak, and I'm a medical officer in the Molecular Genetics Branch here. We have eight speakers who have volunteered to present today, and each speaker has been allotted 4 minutes. Our first speaker is Dr. Dara Aisner of the University of Colorado, and let's get started.

DR. AISNER: Thank you for this second opportunity to speak. I want to discuss today that there are many core elements which are critical for high quality testing. In my mind, these are no different for next-generation sequencing compared to other platforms because NGS is a platform like so many others. So I'd like to highlight what I think are some of the core elements for quality testing, which can be applied to NGS or other platforms.

The first of these is validation. While validation remains a core component of quality, it's far from the only component of core quality. Particularly, when I think about it in the context of next-generation sequencing, validation has limitations.

As has been discussed today, there is simply no such thing as reference materials which can

encompass every possible alteration for every possible assay design. Particularly when we acknowledge that the ability to detect an alteration is dependent on assay design, the location of key reagents such as primers and probes, those things dramatically impact what we can detect and the analytic sensitivity at which we can detect it.

So pictured here is an example of how validation might not identify key flaws in design. In this instance, we see KIT exon 11, and this is a classic deletion for KIT exon 11 in gastrointestinal stromal tumor. These primers have been optimally designed for this specific deletion. However, there is this rare deletion that would not be detected due to the assay design. This is something that nobody would be able to predict unless they happened to have that sample on hand.

So again, this is about the fact that you really can't have a sample depth that you can know what's in every sample. So validation by itself does not give you the full picture of what the assay can and cannot do.

Ongoing quality control in a laboratory is critical. This can be seen in the use of appropriate QC tools, controls, metrics, repeat testing, orthogonal testing, and all of these are critical measures for ongoing quality management in a laboratory. Importantly, proficiency testing becomes extremely important here. Proficiency testing allows you to address the question of you don't know what you don't know.

All of these QC efforts are completely meaningless without the flexibility to act upon them. Laboratories can and do identify weaknesses be they in design or process. The ability to identify and then accommodate these identified weaknesses is perhaps one of the most important components of maintaining high fidelity testing.

Assays which are locked in are ones in which identified deficiencies cannot be corrected and can lead to patient harm. By locking in assays to the point that we cannot apply science to improve them implicitly states that we think that the test is more meaningful than the biology. Which matters more, the

test or the biology? I would argue that we should be keeping the biology in our sites and recognize that testing is a means to understand it and not override it.

As a practicing molecular pathology laboratory physician, I can state unequivocally that one of the most important elements for me to provide quality care to my patients is the ability to see the data. I demonstrated this once earlier today. I'll show you this again with a non-NGS example.

This is a real-time PCR assay for EGFR. This is the same assay that is approved for FDA use, however, we run it in a non-FDA approved manner so that we can see the data. In this particular example, this curve here for an EGFR alteration is out of range. If this assay had been run in FDA-approved mode where we would not have seen the data, this would have been classified as negative. The patient would have gone on to standard chemotherapy, and nobody would have been any the wiser.

However, because we were able to view the data in this case correlated with the percentage tumor,

we were able to alert the clinicians that we thought there was a high probability that there was actually an EGFR mutation in this case, which there was, and the patient is now on therapy and responding. So this is but one of many examples.

Lastly, I would like to emphasize that molecular testing, no matter how simplistic we may make it seem through pre-packaged kits and instruction, should always involve the expertise of appropriately trained personnel. I would like to think we can reduce these systems to a sample-in/answer-out basis. The reality is that these issues are much more complex. This isn't actually even my more complicated of these flowcharts.

So you can see that we have this very complicated flowchart and we cannot equate this to a sample-in/answer-out type system. Thank you very much.

DR. PATHAK: Thank you for your perspective, Dr. Aisner.

Our next speaker is Dr. Toby Guennel from Precision for Medicine.

DR. GUENNEL: Thank you. I'm Toby Guennel,

director of analytics at Precision for Medicine, and we are supporting our clients and sponsors in going through the FDA submission process. Specifically, my role is usually the role of statistics and support, which means I'm usually the bearer of bad news that has to do with two worlds and sample size.

What I wanted to talk about today is not so much providing solutions to how we can use statistics to alleviate some of the burden, but really challenge the statisticians and bioinformaticians in the room to do whatever we can do to push this space forward.

Just as a motivating example, traditionally, some of the studies that involve clinical specimens and analytic validation studies may be LoD, precision, accuracy, clinical validation studies say for a CDx follow-on claim and maybe for a methods comparison study. These are the four studies that usually require quite a bit of clinical specimens.

Usually then, I get asked, well, what is the sample size that you would need to support these studies in terms of clinical specimen, and let's assume we have a panel for non-small cell lung cancer, and my

answer then would be, well, we need a 100 ALK positive specimens across all of these studies. That's usually when it gets very quiet in the room, and everybody wishes that they wouldn't talk to statisticians ever again.

Why is that? Well, the prevalence of ALK is only 1 percent, so in order to actually get 100 ALK positive patients, you would have to screen over 1,000 patients, given that you have to first find patients where you have sufficient tissue to run multiple tests. You have to make sure it's representative of the intended use populations. And you may have to run multiple reference methods due to the sensitivity issues.

So what can we do to think about how we can use statistics and bioinformatic and to really evaluate the performance of a test? Let's assume we have a nicely developed test. We have shown commutability between clinical and contrived samples, and now we want to use some statistical approaches to evaluate what performance metrics are impacting the performance of the test.

Assume we have heard in the earlier panels that you can establish where you can use in silico approaches to evaluate and disrupt the system and introduce diversity. So we can introduce spike-ins. We have simulation studies to evaluate numerically the impact of certain parameters on clinical and analytic performances.

In terms of clinical validation studies, the question may be slightly different. Assume we have a CDx that has established clinical validity with a clinical outcome -- using clinical outcome data. Now the question is, if we can show a certain level of performance between an NGS panel and a companion diagnostic, what can we do to then establish clinical validity for an NGS panel?

We can use simulation studies potentially to evaluate the impact of different parameters on performance, for example, and concordance between the CDx and NGS panel, but also, a lot of the other parameters that claim to play a role in actually establishing clinical validity. We may be able to use adaptive designs to show as an alternative to establish

clinical validity in more of a phased approach rather than having to procure a lot of clinical samples in a short amount of time. And finally, we may be able to use simulation studies to evaluate the impact of using contrived samples in the establishment of clinical validity.

So as a summary, for a single submission, a lot of times it may be necessary to screen thousands of samples, and this can be a very challenging approach to get an NGS oncology panel approved. Alternative approaches to the traditional testing paradigms are needed, and the question in this discussion paper shows that a lot of progress has been made in the recent months.

The richness of the data that the NGS panels are providing can be leveraged to evaluate analytical validity as well as clinical validity, and we should consider in silico approach as a viable supplemental approach to evaluate the impact of parameters on performance for both analytical and clinical validity.

DR. PATHAK: Thank you, Dr. Guennel.

Our next speaker is Dr. James Willey from the

University of Toledo.

DR. WILLEY: Thank you. The data I'm presenting pertain to the use of synthetic internal standards as process controls to establish performance characteristics for each NGS-based rare variant test. I'm focusing on three performance characteristics that contribute to measurement confidence and the use of synthetic internal standards to control for them. We then incorporate this information in the analytical pipeline to assess the confidence for each measurement.

The first source of variance is stochastic sampling. When loading sample into the library preparation, we assess this by making a mixture of two cell lines homozygous for opposite alleles at the same site, made in 1 to 1 mixture, and then made extreme limiting dilutions of one cell line relative to the other. We then mix those dilutions with a known number of internal standard molecules. I am presenting data just for this one SNP, although we can do this for hundreds of targets.

As is clear, where there's a limiting number of molecules loaded, when you control for a high

sequencing count, as the number of copies are loaded, the 1 to 1 measurement, even a 50 percent allele representation becomes unreliably measured, as one would expect.

We then also did a serial dilution of the libraries generated, and when controlling for the number of copies loaded and simply looking at the sequencing count, again going down to low sequence copies, the measurement of even a 1 to 1, 50 percent representation become unreliable. The point is that each of these stochastic sampling events independently contributes to lack of confidence in measurement. So inadequate loading at each step is important to control for.

The other component that we're using internal standards to control for is the sequence variation at each actionable mutation site. As it's known, the sequencing error rate varies from one nucleotide to another. It can also vary from region to region as well as the position within an amplicon that's being sequenced.

What is clear here is that, as is known,

different transitions and transversions have different error rates. Importantly, the internal standard replicated the error rate in the targets extremely well, and the variation that you see here is largely regional dependent. We control for that by looking at each nucleotide in the internal standard relative to the native template and see a very strong correlation. And any deviation from that line is largely explained by the stochastic sampling from the number of counts or number of molecules loaded.

So once one does control for the sequencing counts and the copies loaded, there's a high degree of confidence about what the sequencing error is for a particular nucleotide.

As an example of amplification of this information, this is an analysis of the KRAS G12D mutation. As you can see as you load lower numbers of molecules with the orange -- or red I guess here -- as the mutant molecule's copy is loaded, and the purple as the number of sequencing counts, you can see that you may have a high sequencing count, for example, for this one around 300. But because of the low number of

copies loaded, the CV is very high, and one should not rely only on the number of sequences.

The point is, it is possible to measure the number of copies loaded for each test. It is possible to measure the number of sequencing counts for each test. Each of these should be measured and contributed to the level of confidence around a particular measurement.

The conclusion is, then, CV should be estimated for each variant fraction measurement. It is easy to do. And the method I presented is probably not the only one, but it can be done with this method. And based on the molecules loaded in the library and the library amplicons measured into the sequencer, synthetic internal standards in each measurement as process controls is an efficient way to estimate CV for each value and sequencing error at each nucleotide.

Any departure from optimal conditions will be associated with higher LoD, whether that's loading small samples from cyto prep, from FFPE that reduces the number of measurable molecules, any of these things contribute. Suboptimal conditions are frequent,

unpredictable, and can render a 5 percent measurement unreliable. And this can be due to, as I said, the sample quality, the size of the sample, reagents, library preparation.

Under optimal conditions, which I would say here is feasible, is 50,000 amplifiable copies loaded into the library with 1,000 library amplicons sequenced. The limit of quantification we measure in our particular conditions for a KRAS G12D mutation would be about 0.4 percent assuming 200 mutated copies, 50,000 wild type copies, 1,000 sequences measured for each value. This will be associated with a CV of about 20 percent and a 0.2 percent sequencing error at this site.

So LoD defined as 3 sigma above the background would be quantifiable for each measurement. We incorporate this information into our pipeline for each test, each measurement, and each sample, and I think it's doable. Thank you.

DR. PATHAK: Thank you, Dr. Willey.

Our next speaker is Dr. John Sninsky from CareDx.

DR. SNINSKY: Good afternoon. I'm John Sninsky, CSO with CareDx. I want to thank the FDA for their efforts in organizing this workshop and for the opportunity to share CareDx's recommendations in regard to these panels.

Organ transplant patients are at significantly elevated cancer risks due to chronic immunosuppressive therapy. Indeed, younger organ transplant patients, as indicated here, are at even higher risks than older patients. Lastly, organ transplant patients who develop cancer had been reported to experience worse outcomes than patients with cancer in the general population.

CareDx is a molecular diagnostics company focused on the discovery, development, and commercialization of clinically differentiated, high value, personalized diagnostic surveillance solutions for transplant patients. CareDx brings substantial experience from context to questions concerning clinical diagnostic testing. We currently serve the 2500 heart transplant patients that receive new hearts each year and for subsequent years following their

transplant.

In 2005, we launched AlloMap, a gene expression test from our CLIA-CAP certified laboratory. Crucial CareDx supported evidentiary clinical trials, and more recently registries, demonstrate that the AlloMap solution aids in the identification of patients with stable allograft function.

In 2008, we received FDA clearance for AlloMap, and as a result, AlloMap is one of only a handful of laboratory developed tests that are both CLIA-CAP certified as well as FDA cleared. We have tests at approximately 70,000 samples from heart transplant recipients since 2005. Further, more recently we released an analytically validated cell-free DNA assay using next-gen sequencing to monitor donated organ injury.

This FDA workshop requests input on the establishment of analytical performance characteristics of oncology panels and the production of clinical information needed to support following companion diagnostic devices. To that end, we offer the following recommendations based on our experience with

advanced diagnostic testing in NGS.

We recommend the FDA use fit-for-purpose design concept standards for tests using NGS technology by testing at boundary conditions to ensure proper design and validation. For those DNA variants shown not to be sufficiently validated by these standards alone, we recommend that the FDA work with the manufacturers and test developers to establish predefined individual performance standards.

Clinical grade and research grade NGS testing are very different. This slide summarizes some of the important differences between these two types of NGS tests. But due to the allotted time, it doesn't permit me to comment on them separately. However, I would like to emphasize the bottom row, which is the importance of well curated, frequently updated databases, and would encourage the FDA, like CFTR, to ensure the presence of such databases.

Sustainable well characterized reference standards to evaluate assays employing NGS are critical. For example, Horizon Discovery has developed digital PCR characterized reference materials comprised

of cell lines engineered to harbor key oncology variants to mimic clinical samples. These standards have allowed us to demonstrate the accuracy of oncology panels from both Qiagen and Illumina as noted on this slide.

Standardization of computational analysis is paramount. We applaud the precision FDA initiative to explore and encourage cloud-based analysis. We further recommend the use of in silico constructed standards blended from sequence reads from different sources. Our extensive experience streaming data from our MiSeq sequencers with a secure and auditable DNAnexus platform gives us a high level of confidence that this approach will serve as a powerful tool for community evaluation and regulatory oversight.

It is essential to demonstrate concordance between computational pipelines. We have observed that even minor differences in software modules of pipelines may generate different overall results. This slide denotes near identity of ratios of donor drive, cell-free DNA computed by an analysis on a CareDx local cluster and a pipeline we assembled in the DNAnexus

cloud.

In summary, my colleagues at CareDx and I encourage the FDA and our industry to note the topics on this slide. But due to the time, I want to emphasize just three. First, the importance of using clinical grade sequencing procedures; secondly, to continue to iteratively review high confidence regions and improve low confidence regions of the NIST reference genome, and lastly, to identify flexible and adaptable regulatory approaches to address the dynamic accumulation of evidence.

All of us owe it to the patients we serve to provide results with which we have the utmost confidence.

DR. PATHAK: Thank you, Dr. Sninsky.

Our next speaker is Dr. Roger Klein from the Association for Molecular Pathology.

DR. KLEIN: Thank you. I'm a professional relations chair at Association for Molecular Pathology, and I oversee molecular pathology at Cleveland Clinic. Thank you for the opportunity to present.

We believe at AMP that there is a need to

modernize FDA regulation of in vitro diagnostic test kits that are manufactured and sold to laboratories. This is includes the development of more consistent and predictable regulatory pathways with reasonable requirements that are appropriate to the context in which a test is generally used. Thus, our remarks should be solely viewed in reference to FDA oversight of instrumentation and reagents.

FDA appears to recognize the decline of the single biomarker, single drug, single test paradigm. However, the content of today's discussion and the white paper that was released last week suggests that the agency may fail to appreciate the present unworkability of the companion diagnostic approach and its fundamental incompatibility with the use of massively parallel sequencing technologies for oncology applications.

FDA seeks information about and dichotomizes putative requirements for establishing analytical validity for variants used as companion diagnostics and then other variants used to guide treatment in patients who have "exhaustive therapeutic options." However,

from the standpoint of the analytical performance of the sequencing instruments and reagents, these two applications should not differ at all.

Rather than inserting itself into medical practice, any FDA efforts in next-generation sequencing for oncology should be directed toward ensuring that instruments, informatics, and reagents perform accurately and reliably. This validation should be method based, should primarily entail proving the reliability and reproducibility of the range of mutations likely to be representative of those faced in clinical settings, and should include a representative sample of specimen matrices with which the tests will likely be utilized.

Although an intended use statement that sets forth the generic idea of using sequencing to detect and characterize nucleic acid sequence variations is appropriate, an intended use statement that incorporates the concept of companion diagnostics is not.

FDA's proposed intended use language illustrates the agency's continued adherence to an

obsolete companion diagnostic test approach that is inconsistent with the nature of massively parallel sequencing technologies and their current and future capabilities. The medical use of this powerful, broadly functioning instrumentation is and will always be within the discretion of patients' physicians, including pathologists, the latter of which are responsible for interpreting NGS cancer panels.

FDA must recognize that the clinical interpretation of accurately called bases is an integral part of professional and medical practice and does not lie within the agency's purview. Instead, FDA should focus its limited resources on establishing means to ensure that instrument and reagent manufacturers achieve accurate and reproducible interrogation of the genes or gene region sequenced.

Variant assessment requires knowledge of the performance characteristics of the assay used, but is also dependent on key intrinsic specimen parameters, such as the allele proportion of a variant, which itself is dependent on the proportion of tumor cells in a sample, the zygosity of a mutation, and other

features such as copy number, aneuploidy, and chromosome loss.

Finally and critically, contemporary review of the medical literature is required. Both the analysis and assessment of these complex variables and the specifics of particular variant interpretation entail the use of considerable medical judgment. These activities are the work of pathologists and other laboratory professionals who perform NGS tumor testing. Most important, they are well beyond the scope of FDA's expertise and the agency's ability to positively contribute to patient outcomes.

We heard a patient representative today talking about interpretation of tests. We are the people who do that interpretation. That interpretation is a professional function. Thank you very much.

DR. PATHAK: Thank you for your perspective, Dr. Klein.

Our next speaker is Dr. Garlick Russell from SeraCare Life Sciences.

DR. RUSSELL: Thank you to the FDA for this opportunity to speak this afternoon. One message late

in the afternoon is that highly multiplex controls for NGS assays. So let me get it over quickly. We've got some data here I want to show you. We've got a terrific InterLab study going, and the members are NCI Mocha, Dartmouth Hitchcock, Weill Cornell, Virginia Commonwealth, BioReference Labs, Jackson Labs, and SeraCare, and Beta Innovations.

Right to the data. We are doing QC standards reference materials. That's what SeraCare is all about. Earlier in the day, we talked about pre-analytical validation in QC. All I want to say about that is extracted DNA from the patient's sample is the internal control. So what I'm going to talk about is from library prep on, right through bioinformatics.

For sequencing and pipeline validation QC, highly multiplex assays require highly multiplex reference materials. That's the message I want to leave with you. It allows you to pool your results to increase your sample size and apply the appropriate statistics. And by doing that, you can also use precise analytical frequencies required to trend data, which we always want to do to make sure the assay's

performing properly and also challenge limit of detection. The best way to know if you can detect a variant is to test it, and you can add those through biosynthetics.

The InterLab study, this is technology obtained through the agreement with the National Cancer Institute. The materials we're using are multiplex plasmids. In each run, we have 52 detectable variants, and let me tell you how we do that.

There are 26 1,000 base pair biosynthetics, each containing a variant, either an SNV in homopolymer, indel, or an SNV in a background genomic DNA, very well characterizes the GM24385 at two different allele frequencies. And then each plasmid includes an actionable variant in a unique 6 base pair internal quality marker. We use digital PCR as an orthogonal method to measure the allele frequency. And this was tested at six CLIA laboratories and one research laboratory, and lots of replicates as you can see over an eight-week study.

We're doing some ongoing data analysis right now. I'm just going to show you some preliminary

results to wet your appetite. We are using non-parametric analysis for the comparisons, and also looking at outliers to see if an assay is actually performing to specification using binomial distribution.

Here are some box plots with whiskers to compare platforms. As you look at this chart, on the Y-axis is the percent allele frequency. The dashed black line is the measured value for that control used in the library prep, and it's measured by digital PCR. On the left, it's 18 percent, and on the right, it is 10.3 percent.

We're looking at the laboratories grouped by platform, either on the green on the Illumina side, or in the red, the Ion Torrent. And as you can see, we get very close results, very good results, just over 1200 data points for each one of those whisker and box plots. And as you can see, there's distribution of outliers as well.

So this is a nice way to compare platform, and the data we're actually very pleased with. I think we're getting very good results lab to lab using this

particular control. And as you can see, on the 10 percent shot on the right, very similar results. The Illumina is actually exactly on for the DPCR as the orthogonal method, and a little bit higher with the Ion Torrent. The Ion Torrent's about 2 percent higher allele frequency on both examples. We can also compare platforms lab to lab types of variants, which we're doing right now, different formats, limits of detection, and of course pipeline analysis.

Another way to look at it is InterLab using the same platform, again, the same type of charts, digital PCR on the left at 18 percent and on the right at 10 percent. And again, you can get very similar results.

Just quickly going to the binomial, when you're running assays, you want to know if that particular assay is in or out of specification. So when you use highly multiplex assays, you can do things like binomial distribution. And again, visually for an audience today, just late in the afternoon, you can see there are differences between the laboratories. We expect 6 percent outliers, and we're getting as high as

10.1 percent on the upper panel -- that's the Ion Torrent -- and 21.2 percent on the lower panel.

Let me just -- recommendations for analytical and sequencing validation performance. First, use of highly multiplex controls where you have more than 50 variants in a particular run as reference material, allows you to pool the data and really do good comparison performance. It gives you a greater chance to detect true assay variability with large data sets.

By using outlier testing -- for example, I just showed you the binomial distribution to complement other trending reports like Levey-Jennings -- very important to accept or reject a run. As we develop NGS, we need new QC methods that are appropriate for these types of technologies.

We've been very happy using digital PCR as an orthogonal method to test the allele frequency, to test it against the next-gen sequencing platforms. Controls and calibration also have a flexible design, so it's easy to add new variants. And these are, as I said, biosynthetic plasmids. It's easy for us to manufacture these in hundreds of different variants in a particular

mix. Thank you very much.

(Applause.)

DR. PATHAK: Thank you, Dr. Russell.

Next, we have Dr. Natalie LaFranzo from Horizon Discovery.

DR. LaFRANZO: Good afternoon, and thank you to the FDA for organizing this workshop and for supporting this open discussion. I'm here representing Horizon Diagnostics, a division of Horizon Discovery. Horizon is a translational genomics company with a suite of gene engineering technologies. A key area of our focus is providing genetically defined reference standards to help laboratories develop, validate, and monitor their assay performance. Our standards have been used in about a thousand laboratories worldwide and in a variety of molecular assays.

Relevant to today's discussion, we are also trusted partners for companies who are developing NGS-based companion diagnostic assays, where we provide reference materials to identify and minimize the sources of variability in these assays, from DNA extraction through bioinformatics analysis.

At Horizon, our expertise is in engineering human cell lines, and this provides the backbone for our reference standards. Starting from a standard human cell line, we perform a single cell dilution to access a clonal cell population. We characterize this cell line, and then we begin engineering using the most appropriate gene editing strategy. Once generated, we perform additional validation and present the materials in the appropriate format, whether that's genomic DNA, formalin compromised DNA, or formalin fixed and paraffin-embedded cell lines ready for extraction.

This flexibility and format type provides assay developers the opportunity to develop pre-analytical QC checkpoints, which we believe are essential to monitoring performance.

We believe cell line drive materials have significant advantages as reference materials for NGS-based oncology assays. First, they offer the option of generating highly customized materials, which mimic clinical archives and disease states. Because our variants are engineered directly into human cell lines, they are presented in the appropriate genomic

context with all the complexity and diversity that is included in a true genome example. This approach allows us to access specific parameters that are important to evaluate for NGS-based oncology assays, including variant type, variant size, and both a local and global sequence context.

With engineered cell lines, large structural variants and copy number variants in their true genomic context can be realized. Using mixtures of isogenic mutant and wild type cell lines, we can provide materials at a broad range of allelic frequencies tailored to the appropriate assay or disease state. This provides an easy means for determining the limit of detection of an assay, which we have discussed as an essential step in the assay development.

Finally, the ability to multiplex desired variants into a single sample enables the use of a reference material to be affordable for routine use.

At Horizon, we feel strongly that good manufacturing is the key to providing renewable, trustworthy reference materials for the development, validation, and routine monitoring of companion

diagnostic assays. While clinical archives offer diversity in sample types, our manufacturing process enables the same sample to be used across many sites and over many months or even years.

Small clinical labs running NGS-based oncology panels will likely not readily have access to cell culture facilities or large biobanks of patient materials, but these labs still require suite or reference materials. And we believe routine monitoring is essential for NGS-based oncology panels, especially when we consider the run-to-run variability introduced by region batches, operators, sample handling, et cetera. And the data presented here, collected by our clients at St. John of God pathology, highlights the run-to-run variability of a commercially available oncology panel collected over a period of eight months.

The same reference standard with 11 verified cancer mutations was extracted from an FFPE curl routinely sequenced and analyzed, which allowed the laboratory to establish a baseline and appropriate thresholds for their acceptance criteria. This method of routine monitoring can be continued indefinitely,

providing quantitative checkpoints as molecular informatics updates are introduced.

So as we develop new cell line based reference materials, we wish to engage in active discussions with assay developers, with clinicians, and with regulatory agencies to ensure our materials are both accessible and tailored to fit the needs of the end user. Horizon has a diverse set of tools readily customized to benchmark the performance of NGS-based oncology panels, and we are eager to share these with the community.

We feel strongly that consistent materials prepared in variable formats, which allow for the effects for evaluation of the effects of sample handling and processing, such as formalin fixation, extraction, and nucleic acid quality, and universal materials, which allow for cross-platform validation, are absolutely essential to evaluate assay performance.

I look forward to hearing your feedback on how we can contribute to these efforts. Thank you.

DR. PATHAK: Thank you, Dr. LaFranzo.

Our final speaker is Dr. Daryl Pritchard from

Personalized Medicine Coalition.

DR. PRITCHARD: I guess I'm the last speaker of the day before the wrap-up.

Good afternoon, everybody. My name is Daryl Pritchard. I am the vice president of science policy at the Personalized Medicine Coalition, PMC. PMC represents over 225 members, including innovators, scientists, patients, providers, and payers. PMC promotes the understanding of personalized medicine concepts, the adoption of personalized medicine services and products, and the advancement of novel cutting-edge technologies that can benefit patients and the health system. We thank the FDA for the opportunity to speak here today.

Next-generation sequencing based oncology panels hold great promise for advancing personalized medicine in cancer care. My comments will be about the regulatory implications of the discussion today. NGS will allow for the identification of all the genetic variants an individual or their tumor can have. This information can be used to make healthcare decisions based on the molecular characteristics of each

individual patient's disease, thus truly personalizing healthcare strategies.

Personalized medicine in oncology relies on accurate and reliable determination of clinically relevant individual patient information. Yet, future investment in technological advancement depends on clear, predictable guidelines. The oversight of NGS-based oncology panels needs to appropriately foster innovation and allow timely access to new personalized medicine information while ensuring accuracy and reliability. This presents both challenges and opportunities.

NGS-based oncology panel oversight represents a new frontier in regulatory processes. Traditional companion diagnostic assays assess a single analyte or prespecified mutations associated with therapeutic response. However, an NGS panel may identify multiple genetic variants concurrently, and the results of the tests could lead to useful information about many different biomarkers related to cancer.

Traditional methods of oversight for diagnostic tests may be obsolete for these kind of

dynamic tools. In devising regulatory oversight schemes, FDA must seize the opportunity to provide meaningful oversight of novel technical diagnostic platforms and dynamic clinical evaluative capabilities.

It is impractical to verify the analytic performance characteristics of every possible genetic variant that could be detected in a genomic sequence, as mentioned here today a number of times. Thus, it is important that guidelines for the assessment of NGS analytical test performance be a clear, predictable, scientific subset of standards based approach in the appropriate sequencing context.

New paradigms and oversight of clinical performance must allow for flexibility in the analysis of information. While the NGS tested cell might not change the general intended uses of the panel, the biomarkers included in clinical assessment and the clinical application of the information provided by the panel may change over time with additional knowledge about individual biomarkers and their clinical significance.

We recognize that this can be a challenge.

For example, there are many different mutations, the EGFR, BRAF, and KRAS genes, related to solid tumor mutation analysis, and NGS has the ability to identify all of the variants in these genes, even as our scientific knowledge of specific mutations expands.

While we cannot assume that the detection of additional mutations provides newer improved clinically actionable information, FDA should describe and then be willing to accept what is considered a reasonable level of evidence to explore additional or expanded purposes.

Additionally, while NGS-based oncology panels may report on variants over a spectrum of clinical claims, it is important to determine an appropriate level of clinical evidence to provide a reasonable assurance of safety and accuracy for included variants. Important factors in clinical decision-making such as risk level associated with a biomarker, whether it is a rare or coexistent variant, and whether there is conflicting data regarding a clinical claim should be considered.

Moving forward, we recommend that FDA continue developing discussion drafts or draft guidance

documents describing current thinking on these issues and incorporating feedback from this workshop and other forums. FDA draft documents should be updated regularly to reflect evolving though and best practices. We urge the agency to take the time necessary to get this right. Future investment in the field depends on clear, reasonable guidelines which are in our power to develop now.

We appreciate the high level of engagement that the agency is having with stakeholders for considering potential new oversight processes, and we encourage FDA to continue this engagement and look forward to working with the agency going forward. Thank you.

(Applause.)

DR. PATHAK: Thank you, Dr. Pritchard.

I'd like to thank all the public comment speakers for their perspectives and expertise. This concludes the public comment section of this workshop.

Next, Dr. Reena Philip, the director of the Division of Molecular Genetics and Pathology, will present a summary and wrap-up of today's workshop on

NGS oncopanels.

Summary and Wrap Up

DR. PHILIP: Thank you all for staying until the end of the workshop. I'll try to provide a wrap-up summary from what we heard today. Definitely, though, video archives will be posted next week, and the transcript will be available, too. Of course, our discussion materials are on the website, and please comment on our discussion materials prior to March 28th.

We had three great panels. The first one was on the pre-analytical and quality metric approaches. We heard that these pre-analytical aspects can have a significant impact on the assay performance, for example, in the sensitivity of the assay, so standards and guidances are necessary. If variations are used in these pre-analytical aspects, then you should consider appropriately reporting them to reflect -- the necessity of such a variation.

Good pre-analytical steps can lead to increased downstream quality of the NGS data. Each of these steps from sample quality to variant calling

should be considered. As we heard, there are several steps in the pre-analytical process starting from sampling decision, selection of the tissues; the regions to be tested in the practice of medicine.

Sample processing has major steps, including gross tissue processing, sample metrics, tumor content enrichment, and nucleic acid extraction. It appears that there are some thoughts that the assay will start from the nucleic acid step, so manufacturers should optimize different nucleic acid isolation methods and provide that data when they submit the submission.

Then we heard about tumor content, tumor cellularity. Those aspects are very important. Tumor cellularity requirement varies depending on if it is for SNV versus LOH detection. Again, how enriched the tumor is and what is the estimated tumor content is still quite subjective. We also heard about the internal reference control. Also, different library preparation techniques to be used to test the best assay performance.

But we heard that the most important metrics are the post-sequencing QC metrics. There were three

different criteria: percentage of reads mapped, percentage of reads on target, percentage of unique reads, which reflect the library complexity, which we heard is really important. Coverage is also very important; so uniformity of coverage is another important metric.

We heard that the technology versus platform, how it's different; the hybrid capture versus PCR based amplicons capture are different – even the QC parameters for the approaches are different. Also, the first panelists touched upon the contamination, and we heard to look for the contamination and also for -- there may be specimen mix-ups, maybe sequence tumor, tumor normal pairs.

Talking about the percentage of cancers, whether that can be used to demonstrate Pan-Cancer claims. We heard difficult and challenging tumor types are important, for example, bone, and tumors presenting potential source of assay interference and presenting different matrices should be assessed; example, mucin, necrotic tissue. I think we usually ask for those, so we heard from the panelists, which was encouraging

actually.

Then regarding using clinical samples, we heard large numbers of clinical samples should be used. There were some different opinions, numbers, but we did hear that there were -- and it's good to have clinical samples, should be used spanning several variant types; in the case of rare variants, of course there could be exceptions.

Then we had the very exciting second panel, which is the analytical validation, bioinformatics, and post-approval assay modifications. Regarding the representative variant approach, the number of samples are as important as validation of all possible mutation types and mutations with different genomic contexts.

We also heard the need to include number of samples to cover different types of variants. It also depends on the intended use of the assay. One should be assessing all different variant types that are included in the intended use; that is CoDx and non-CoDx. The level of analytical validation is variable for the different he variant types we heard about. And all the different types should be tested

for parameters like sensitivity, specificity, accuracy, reproducibility.

Then there was a question about indels should be considered in the different -- the ranges. And the panel did say these questions are really challenging. But the design of the assay is important. In silico approaches are acceptable, but one should go back to biological assay to determine the sensitivity and specificity, matrix of the assay.

We heard the whole genome should be assessed and not just the hot spots. More reads are necessary to confidently call indels. We heard about the depth of sequencing, surrounding sequence complexity, and the cutoff by which a variant can be detected with confidence is important.

Regarding the LoD, we heard that LoD has to be variant specific for broader claim; should be shown that a specific claim can be generalized to the new variation. We also heard variations in the labs could lead to differences in the NGS data. There were questions regarding the acceptable orthogonal method for the accuracy, and we heard alternate NGS approach

is acceptable. But the new data sets should be compared to some samples with truth variant calls.

Sensitivity and specificity overall are based on target regions. It's important. And then there was discussion on the somatic versus germline variants. Many labs are using tumor samples solely for detection of somatic variants. But in general, the most confident call is based on comparison with matched normals. Alternatively, there should be strategies to filter out the germline variants in somatic-only panels.

There was some discussion which we wanted to hear more about the modification of approved panels. The panelists said any modification should be validated. It may not be the same as the original validation, but key performance metric should be evaluated following the modification. And the device manufacturer has to test any modification -- has no negative impact on the other members.

Then moving on to the third panel, clinical and follow-on companion diagnostics claims, we heard for the follow-on companion diagnostics, the pre-

analytical issue is important. What kind of specimen is used? If the original study had the FFPE, it's expected that the follow-on also uses FFPE. Regarding the matrices effects, the interfering substances should be assessed and trials should be designed, such that labs should be using the same technology. There were also some comment that companies who sold the tests should provide the control material and positive control testing should be done.

There was some discussion on after the FDA approves a panel, the labs that are getting this panel, some labs have more experience, and the other labs may have very limited experience. That's where this follow on is important. Is it just to detect the variant is important or should extensive validation. But I think we heard -- the thinking from the panel is very along the lines of what we have been asking.

We also heard if it's a follow-on, reflex to companion diagnostic tests is what people have been doing. We heard a lot about the standardization. There has to be some standards before recommending the treatment. Carefully collect the data and understand

the clinical relevance.

We talked about clinical trials essential to understand what molecules have specific interaction with certain drugs, especially the ones which are the non-companion diagnostics. So prospective data collection is important and standardization methods should be followed to find outcomes for treatment decisions.

Then we heard a lot about transparency, truth in labeling. Be clear what they used. If they used cell lines, that should be there so maybe labs can validate using those cell lines. And we are clear on what informatics pipelines were used, how calls were made; so a lot of transparency should be there. And then the non-companion diagnostics; maybe a database should be created that can be collaborative data sharing can be accomplished across communities.

I think ClinGen's database is like that. It's very important. We heard maybe BRAF and -- there have been melanomas, and it's working, but it may not work in the others. So we have to build up that data to make sure that patients are not treated -- they're

not getting the right treatment.

With regards to sensitivity, I think we heard that if sensitivity differs, it can expose patients to a lot more toxicity if it's really sensitive (a hundred times); is this meaningful. Then there were discussions about how clinicians may not be able to detect and interpret the reports, so how do you support the reporting? How simple should it be? Are there different levels of reporting for patients, for clinicians, for oncologists, and who was the expert in reporting this.

Somebody has to move -- for intended use from the table 2 to table 1, we heard there should be some sort of validation. It may not be as much as what is typically needed for table 1 -- if it's already in table 2, but there should be still some sort of validation.

I believe I've covered everything. With that, I thank all the panelists for actually being here. All three panels were very great, with productive panelists. I also thank the audience, and I also want to thank my team. We have been working for

I don't know how many months -- on this workshop, so I really thank them.

We will use this data, and the feedback is very important for us to evaluate our review as we review the submissions. And we probably will be drafting some guidance incorporating your feedback, so please send your replies to the docket, and we look forward to that.

With that, I think I'll close the meeting. Thank you all.

(Applause.)

(Whereupon, at 4:24 p.m., the meeting was adjourned.)

CERTIFICATE OF REPORTER

I, Janet Evans-Watkins, the officer before whom the foregoing proceedings were taken, do hereby certify that the proceedings were taken by me in stenotype and thereafter reduced to typewriting under my direction; that said proceedings are a true record; that I am neither counsel for, related to, nor employed by any of the parties to the action in these proceedings were taken; and, further, that I am not a relative or employee of any counsel or attorney employed by the parties hereto, nor financially or otherwise interested in the outcome of this action.

_____

JANET EVANS-WATKINS

Notary Public in and for the

State of Maryland

My commission expires: July 8, 2016